



Review

Aptamer Bioinformatics

Andrew B. Kinghorn, Lewis A. Fraser , Shaolin Liang, Simon Chi-Chin Shiu  and Julian A. Tanner *

School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR China; kinghorn@hku.hk (A.B.K.); lewis-fraser@hku.hk (L.A.F.); shaolin2@hku.hk (S.L.); simon156@hku.hk (S.C.-C.S.)

* Correspondence: jatanner@hku.hk; Tel.: +852-3917-9472; Fax: +852-2855-1254

Received: 31 October 2017; Accepted: 20 November 2017; Published: 24 November 2017

Abstract: Aptamers are short nucleic acid sequences capable of specific, high-affinity molecular binding. They are isolated via SELEX (Systematic Evolution of Ligands by Exponential Enrichment), an evolutionary process that involves iterative rounds of selection and amplification before sequencing and aptamer characterization. As aptamers are genetic in nature, bioinformatic approaches have been used to improve both aptamers and their selection. This review will discuss the advancements made in several enclaves of aptamer bioinformatics, including simulation of aptamer selection, fragment-based aptamer design, patterning of libraries, identification of lead aptamers from high-throughput sequencing (HTS) data and in silico aptamer optimization.

Keywords: aptamer; simulation; in silico selection; molecular dynamics; fragment based design; HTS

1. Introduction

Aptamers are short nucleic acid sequences capable of specific, high-affinity molecular binding [1,2]. Aptamers are isolated via SELEX (Systematic Evolution of Ligands by Exponential Enrichment) (Figure 1), an evolutionary process in which successive rounds of selection and amplification are used to enrich an aptamer library for high affinity aptamers. Aptamers are among the simplest of genetic entities, having both genotypic and phenotypic properties and being capable of heredity in an in vitro selection experiment. Their combinatorial complexity poses many questions and problems that are well suited to computational analysis. Many computational approaches have been applied to aptamers, bringing together different disciplines and technologies. This review encompasses a broad range of aptamer bioinformatics approaches including simulation of aptamer selection, aptamer selection by molecular dynamics, patterning of libraries, identification of lead aptamers from high-throughput sequencing (HTS) data, and in silico aptamer optimization. We aim to describe and contrast these methods so that aptamer scientists might make use of the diverse array of bioinformatics resources available.

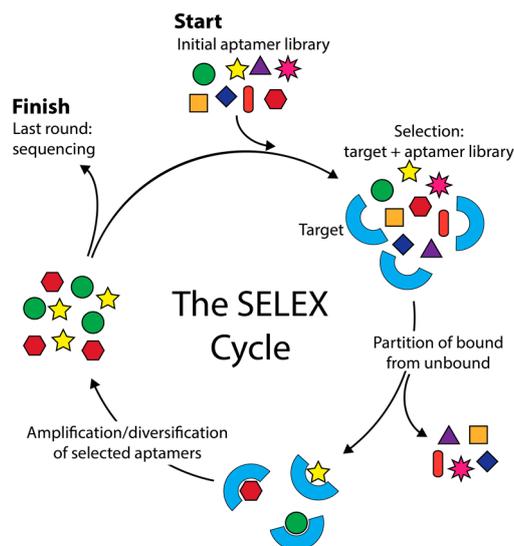
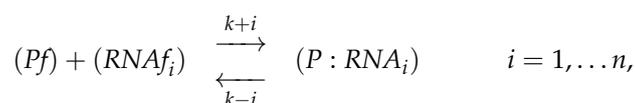


Figure 1. The SELEX (Systematic Evolution of Ligands by Exponential Enrichment) cycle. SELEX starts with a random nucleic acid aptamer library which is used to initiate the SELEX cycle (top arrow entering cycle). The library is incubated with the target and the target is washed to remove and discard unbound aptamers (right arrow exiting cycle) before the bound aptamers are eluted and amplified by PCR. The amplified sequences seed the next round of SELEX. Typically, around 12 SELEX cycles are performed before sequencing and aptamer characterization (left arrow exiting cycle).

2. Simulation of Aptamer Selection

Aptamer selection is complex. Complexity is found in both the myriad of experimental parameters and the combinatorial complexity of nucleic acid libraries. McKeague et al. performed a statistical analysis of 492 SELEX experiments, investigating experimental parameters such as choice of target, selection template, pH, and temperature [3,4]. Specific parameters were shown to have a significant effect on the dissociation constant of the tightest binding aptamers [4]. This information is valuable to aptamer scientists, but is limited to routinely disclosed experimental parameters. Many useful experimental parameters are not routinely disclosed, such as mutation rate, target concentration per selection cycle, recombination techniques and the inclusion of novel unnatural bases. An exhaustive empirical analysis, involving SELEX with sequencing of every round, is limited further by the combinatorial complexity of nucleic acid libraries, which contain $\sim 10^{15}$ sequences in an initial aptamer library pool [5]. Empirical analysis of anything close to this number of library members is simply not feasible. To investigate the experimental parameters of aptamer selection, simulation has been used.

In 1991, SELEX was first simulated using a program named SELEXION (Systematic Evolution of Ligands by Exponential Enrichment with Integrated Optimization by Non-linear Analysis) [6]. SELEXION was first used to reconstruct bacteriophage T4 DNA polymerase gp43 SELEX experiments [2]. A library of eight random RNA bases underwent eight rounds of SELEX comprising ligand binding, partitioning, and amplification. Ligand binding was modeled using a kinetic mechanism between target-protein and all aptamer ligands that reach equilibrium, stated by Irvine et al. [6] as follows:



where (Pf) was the free protein concentration, $(RNAf_i)$ was the free RNA species of i concentration, $(P : RNA_i)$ was the protein-RNA species i complex concentration, $k + i$ was the rate constant for association of free protein and free RNA species i , $k - i$ is the rate constant for dissociation of

protein-RNA species i complexes, $(P: RNA_i)$ was the protein-RNA species i complex concentration, and n is the number of RNA sequences with a unique set of rate constants [6]. Partitioning efficiency for the reconstruction was set to 80% of bound aptamers and 0.1% of unbound aptamers. Amplification of partitioned aptamers involved reverse transcription to cDNA and PCR amplification before library generation using transcription. The experimental parameters of the gp43 selection [2] were reconstructed and underwent simulation as a proof of principle for SELEXION [6]. These simulations indicated that the equilibrium mechanism proposed above for SELEX was sufficient to explain the high levels of enrichment after just a few rounds observed in the laboratory experiments. Following the reconstruction simulation of the gp43 selection, several properties were investigated using SELEXION including predicted enrichment under different conditions, optimal protein concentration when dissociation constant (K_D) estimates are known, near-optimum protein concentration with no estimate for K_D , determination of sufficient protein concentration with no estimate for K_D or background, likelihood of SELEX success, and finally sequence representation in the random library pool [6]. SELEXION took a thorough approach to modeling ligand binding. However, a possible shortcoming would be the determination of aptamer properties such as K_D . The binding affinities in terms of K_D were distributed without reference to aptamer sequence. For the reconstruction there were just five unique K_D values for all 65,536 unique aptamers in the initial library.

In 1998, Irvine et al.'s work [6] was extended and the program MultiSELEXION was coded to investigate SELEX against multiple targets [7]. MultiSELEXION allowed the investigation of problems arising from the use of contaminated protein preparations in SELEX, as well as analysis of complex target selections such as Cell-SELEX [8] and in vivo SELEX [9]. It was found that in most cases SELEX is capable of isolating differing ligands against the different targets in a heterogeneous mixture, irre SELEX spectively of large variations in target concentrations or aptamer/target affinities. However, a low relative partitioning efficiency for a given target in a mixture gives a greatly reduced rate of selection of high-affinity aptamers [7].

Similarly to Irvine et al. [6] and Vant-Hull et al. [7], Chen et al. devised a SELEX simulation model that uses ligand binding based on equilibrium between target aptamer ligands and was applied to subtractive SELEX [10] as well as SELEX against a complex mixture [11]. This difference highlights and simulates selection pressures in SELEX experiments. Further similarities to Irvine et al. [6] and Vant-Hull et al. [7] included the binding affinities in terms of K_D being distributed without reference to aptamer sequence. Chen et al. used just 10 unique K_D values for all aptamers in the simulations [10].

Wang et al. developed a model that focused on the two SELEX parameters, target concentration and the effect of nonspecific binding [12]. The model represented ligand binding using equilibrium kinetics similarly to Irvine et al. [6] and Vant-Hull et al. [7]. Partitioning was modeled in two ways: without background binding, which was intended to mimic microfluidic selection; and with background binding, which was intended to mimic nitrocellulose filter-based separation. Aptamer binding properties were normally distributed [12] as hypothesized in the literature [13,14]. Wang et al. [12] found that "without background binding" conditions, an increasing amount of target decreases the selection efficiency. Under "with background binding" conditions, there is an optimum target concentration that increases with increasing background binding. Interestingly, under multiple selection rounds and "with background binding" condition the optimum target concentration for achieving maximum enrichment increases with each SELEX round. This is contrary to the generally accepted practice of reducing the target concentration as SELEX progresses. The reason for this trend of increasing optimum target concentration in successive SELEX rounds could be the modeling of background binders. The more target, the greater the number of specifically binding aptamers make it to the next round, therefore the higher the ratio of specific to nonspecific binders and the higher the average K_D value. One aspect that SELEX Wang et al.'s model and many other models do not take into account is the possibility of adaptation in the aptamer pool giving rise to aptamers with increasingly tighter K_D values.

Spill et al. developed a model that simulates Capture-SELEX and includes non-covalent ligand–substrate immobilization [15]. Aptamer–target binding was represented using a hybrid approach whereby an equilibrium constant is combined with a stochastic probability model. Following partitioning, the amplification of selected aptamers was simulated. Of particular interest is that the initial library K_D distribution has a dramatic effect on the outcome of the simulation. Additionally, the impact of distribution noise and the downstream effects on the total target concentration were assessed. The use of a stochastic model and Monte Carlo simulation highlighted sensitivity of SELEX to stochastic variation. Twenty very tight binders are capable of outcompeting 10^{15} library members or can be totally lost.

Simulation of SELEX has given insight into how an aptamer scientist might optimize the SELEX protocol. SELEX has both vast complexity in terms of the number of sequences (typically around 10^{15}), and informational complexity associated with each individual aptamer’s sequence, folding, and target binding. The aforementioned simulations have focused on representing the vast complexity of SELEX and neglected the informational complexity of individual aptamer sequences. All binding properties of aptamers are selected randomly or from a distribution with no relevance to the aptamer sequence. For particular questions about SELEX, including the role of adaptation and the occurrence of divergent and convergent evolution, a more thorough binding model is required.

Hoinka et al. coded a program to simulate the aptamer selection process called “AptaSim” [16]. AptaSim aimed at realistically recreating the selection process during SELEX with the intention of investigating the effect of error-prone PCR on aptamer selection. An initial library pool was generated using a first-order Markov Model, previously trained on early SELEX round selection data. The generated aptamer was randomly assigned a copy number and binding affinity within a predefined range. Iterative cycles of capture and amplification were then simulated where the capture probability is related to an aptamer’s copy number and binding affinity, and amplification is subject to a specified probability of mutation. The binding model used attributed aptamer affinities at random without relevance of sequence. Additionally, mutated versions of these aptamers retained the original’s attributed binding affinity. While AptaSim was an important step forward in simulating selection, enrichment and mutation copy number, AptaSim did not appropriately represent heritability or represent binding affinities correlated between related sequences, which is required for the study of SELEX as a genetic system.

Oh et al. used a string matching function as a binding model to simulate aptamer selection [17]. All aptamers were given a target binding score based on their similarity to a given “optimal aptamer” sequence [17]. This model does include heritability and binding correlation between related sequences. As string matching is not computationally demanding, this approach can be used for very large library sizes, which is more representative of aptamer selection. The drawback of string matching is that only close-range epistasis is possible and by using a one “optimal aptamer” model, the landscape is cone-shaped and would not represent a true aptamer binding landscape.

Wedge et al. used Kauffman’s NK model [18] to represent ligand–target binding for the simulation of protein-directed evolution [19], a similar field to aptamer selection. The NK model is a robust mathematical model that serves as an objective function relating genotypic sequences to phenotypic fitnesses that make up a fitness landscape. Using the NK model, strings of informational digits of length N are attributed fitness values equal to the sum of each digit’s interaction with K other digits. In this way, epistatic and pleiotropic interactions can be modeled. The NK model has been used to describe many complex systems such as immunology [20], evolutionary biology [21], and economics [22]. The NK model has also been related to aptamers [23]. In Wedge et al.’s [23] work the ligand properties were determined using an NK model in which binary strings of length $N = 100$ were used with random epistatic interactions varying from $K = 0$ to 10. The initial library size was 40,000 and during each of the 10 selection rounds, 1 to 4000 of the tightest binding ligands were partitioned. Varying degrees of selection pressure (number of ligands selected each round), mutation rate, and crossover (recombination) were tested and it was found that optimal directed evolution (DE) parameters were

strong selection pressure, a high mutation rate, and that crossover is only valuable when epistasis is low to moderate ($K < 5$). While these results are valuable to the field of protein-directed evolution, the simulation did not mimic properties specific to SELEX.

The NK model can effectively represent the target binding of polymeric ligands such as proteins and aptamers. Besides the challenges for biological accuracy in representing base interactions within an aptamer, the classical NK model may have limitations in representing some aspects of biological systems. The NK model's greatest utility is that epistasis can be tuned using the variable K . However, this epistasis is reasonably uniform throughout the sequence. To represent some biological systems, a higher amount of epistasis is desirable. As K increases the fitness landscape tends to become more rugged, to the point where it is too chaotic to allow adaptation, a phenomenon is referred to as the "complexity catastrophe" [24].

To overcome the "complexity catastrophe" and use the NK model to represent gene regulation, Altenberg [25] developed "selective genome growth" in 1995. Selective genome growth is an evolutionary approach that selects epistatic interaction in such a way to create a highly epistatic landscape that is smoother than classic NK landscapes with the same degree of epistatic interaction [25]. Altenberg's selective genome growth NK landscape represents gene regulation very well. However, due to the increasing returns of the selection system, an extremely high pleiotropy is attributed to a handful of digits [26]. This highly aggregated pleiotropy is biologically appropriate and accurate for describing gene regulation. However, as each base in an aptamer has a relatively low number of interactions due to its spatial capacity, the highly aggregated pleiotropy is not biologically representative for base interactions within an aptamer.

To overcome this problem Kinghorn and Tanner recently devised the method "selective phenome growth", which generates fitness landscapes with low aggregated pleiotropy that more appropriately represent aptamer binding [26]. The selective phenome growth process involves phenotypic contributors being added to a genotype/phenotype interaction map sequentially in such a way as to increase the fitness of a selected "fit sequence". In this way, a fitness landscape is built around the selected "fittest sequence". The fitness landscapes obtained were compared to empirical aptamer microarray data and were shown to more accurately represent aptamer ligand binding than other theoretical models (Figure 2) [26]. The selective phenome growth model has not yet been utilized in the simulation of SELEX, only described and validated as a model that more accurately represents aptamer binding.

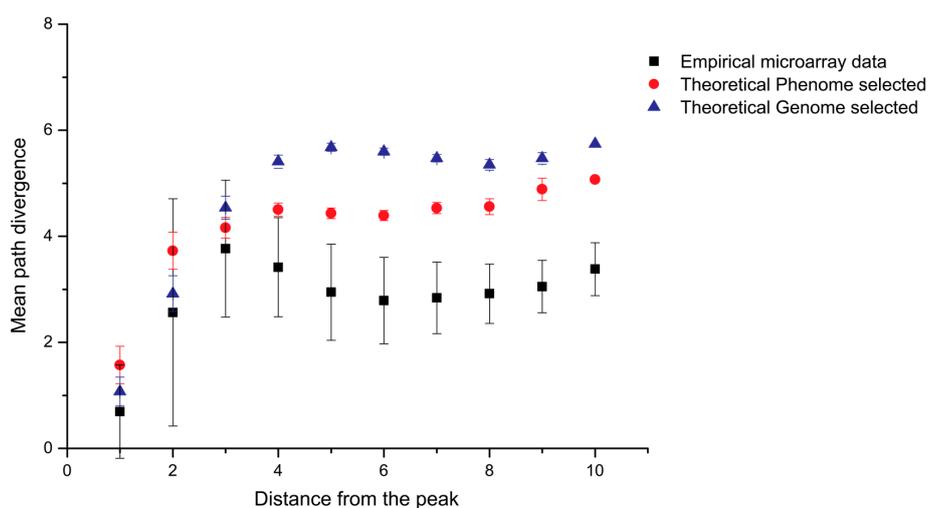


Figure 2. Comparison of novel phenome selected model to both genome selected and empirical aptamer microarray binding data. The mean path divergence analysis, a measure of landscape smoothness, shows that the novel phenome selected landscape is more similar to empirical microarray binding data than the previous genome selected landscape model. Figure adapted from Kinghorn et al. [26].

The no-free-lunch theorem states that all search algorithms perform exactly the same when averaged over all possible problems [27]. This infers that any elevated performance in one class of problem is exactly paid for in the performance of another class of problem. If there is discrepancy between a real-life system and a model used to describe it, for example an empirical SELEX experiment and a SELEX simulation, any elevated performance insight found using the simulation is exactly paid for in the performance of the real-life system. This illustrates the need for the simulation model to be as accurate as possible; otherwise optimizations will not translate to empirical SELEX experiments. The area least accurately modeled in SELEX simulations has been the aptamer binding model.

3. Aptamer Selection by Molecular Dynamics

Molecular dynamics have applications across biotechnology, including but not limited to protein studies, membrane transport, and drug discovery [28–31]. One particular application is to improve the efficacy of aptamer selections by computationally solving the three-dimensional structures of nucleic acids (NAs) and their targets, and simulating the physical forces involved in NA docking to a target. This is achieved by various N-body simulations that calculate the dynamic forces of the atoms and molecules of a NA within a binding site, in the form of a docking score. Docking scores can be used to identify sequences that bind to a target, defining a novel approach for aptamer discovery. Here, we will discuss studies wherein molecular dynamics has been used to enrich selection pools, optimize existing aptamers, and discover new aptamers.

3.1. Whole Aptamer Docking

For the purpose of this review we have divided *in silico* techniques into two categories: those that simulate the molecular dynamics of a whole aptamer and its target, and those that fragment an aptamer into discrete units to simulate binding interactions. We will discuss the literature that underpins *in silico* selections for whole aptamers.

Computationally predicting secondary and tertiary structures of NAs and targets reveals the steric and energetic properties of each structure. These predictions allow researchers to modify their selection pools to have a broader range of three-dimensional structures and NAs with more favorable free energy [32–34], and provides essential information for molecular docking simulations [35]. Many protein-NA structures have been solved experimentally using NMR and X-ray crystallography, for which there are large but limited libraries in the Protein Data Bank (PDB). If the structure has not been solved experimentally, homology modeling webserver services exist for both proteins and NAs [36–39].

In simulating the docking between a target and an aptamer, several non-covalent interactions are assessed including ionic interactions, hydrogen bonds, van der Waal's forces, hydrophobic interactions, base stacking interactions, and shape complementarity [35] (Figure 3A). Algorithms calculate the potential energy between interacting atomic components, known as force fields. For biological systems, the most frequently used MD force field simulations are CHARMM and AMBER [35]. Homology and modeling software for DNA–protein interactions is currently limited [35], as much of the software is based in analyzing protein–protein interactions. A coarse-grained force field has shown how dsDNA interacts with protein structures. Specific interactions are useful but limited in their scope and shape complementarity and internal DNA energy play an important role in simulating protein–DNA docking [40] (Figure 3B).

An initial attempt at *in silico* selection was proposed by Chushak and Stone. Computationally, they decreased RNA sequence search space in a selection pool by up to five orders of magnitude to enable conjugation of an enriched RNA selection pool to a microarray to improve high-throughput aptamer selections [34]. A three-step enrichment approach was used:

(1) Selection based on secondary structure—a set of criteria were used to identify and eliminate sequences with common simple structural motifs and high-energy unstable RNA sequences, both of which would be unlikely to form aptamers.

(2) Selecting for conformational flexibility—a single RNA sequence can have a large range of three-dimensional conformations; the Rosetta RNA package [41] was used to generate these structures. Then the five lowest energy three-dimensional structures, and therefore those with the greatest conformational flexibility, were selected using the AMBER force field simulation and the generalized Born solvation model [42].

(3) Screening the RNA library with computational docking—a modified docking tool called DOVIS using Autodock v4 [43] was used to simulate interactions between all the generated RNA three-dimensional structures and small molecule targets. Docking was scored based on their calculated affinity for the targets. By selecting for the highest scoring sequences they effectively lowered the RNA pool size from $\sim 2.5 \times 10^8$ to 5×10^3 . Six known aptamer–ligand complexes were used to validate this approach. Native aptamers were found in within the top 5% of in silico selected structures.

Confirming that molecular dynamic calculations align with experimental evidence provides further evidence that in silico approaches can complement aptamer selections. A software package that uses the CHARMM force field to analyze protein–protein interactions called Discovery Studio uses a docking simulation algorithm called ZDOCK [35,44]. ZDOCK was found to work effectively with short RNA–protein interactions [45] but was found to be ineffective when simulating longer RNA strands [46,47]. When combined with ZRANK, an algorithm that takes into account a range of attractive and repulsive forces, van der Waal's forces, and desolvation, effective simulation of protein–long-strand RNA was achieved [35,48]. Having confirmed the efficacy of this software package in conjunction with aptamer–protein interactions, Chen's research group mutated aptamers of angiopoietin-2 protein (Ang-2), a protein that regulates angiogenesis and is linked with the development and spread of cancer [49,50]. From the mutated strands, they selected three with high scores and tested them experimentally for binding with surface plasmon resonance (SPR). Based on binding affinity and SPR response, they claim one of these novel aptamers (Seq15_12_35, K_D 0.61) has improved binding when compared to a high-affinity Ang-2 aptamer (Seq1, K_D 1.39) found in the literature [48].

Selection can be a lengthy and costly process [51], especially when targeting human proteins for which native proteins may be expensive or commercially unavailable [52]. To lower the cost of selection, it is common to select an aptamer towards a recombinant or non-human version of the equivalent human protein [53]. There is an increased risk that the difference in homology between the native protein and the recombinant/non-human protein will result in selecting for an aptamer that will not bind to the native protein. This is the case for immune-checkpoint blockade receptor TIM3 [54], for which aptamers selected for murine binding aptamers lacked cross-reactivity with the human form. Based on the murine aptamer, Rabal et al. used a three-step bioinformatics process similar to those already discussed, but coupled cluster analysis with their chosen 3D docking algorithm, 3DRPC [55]. Clustering algorithm GROMACS [56] revealed highly populated clusters focused around specific binding sites. In four out of five cases, combining clustering with docking simulation revealed a binding mode and site that were not identified by docking simulation alone. They were able to show the scope of in silico aptamer–protein analysis by identifying a plausible binding site on murine TIM3 and aptamer binding mode that explains the lack of cross-reactivity in murine over human TIM-3 [54].

3.2. Fragment-Based Aptamer Design and Docking

Whole aptamer selections require massive computational resources and three-dimensional structures of both nucleotides and target. A fragment-based approach has been argued to simplify the process of in silico aptamer generation [57]. Tseng et al. presented a three-step approach in which they only require structural information of the target, known as the entropic fragment-based approach (EFBA) [57]. They first determined the probability distribution of the first nucleotide binding to the target. They then sequentially added nucleotides to the first, taking into account the probability distributions of the added nucleotide to its neighbors and the target. Finally, they determined a cutoff length based on an entropic criterion (information theory entropy). Once the target–NA complex was saturated and the interactions of the complex were at a global minimum irrespective of nucleotide

additions, the sequence was selected [57] (Figure 3C). They developed the in silico “seed and grow” method by selecting two aptamers. One aptamer bound to the target phosphatidylserine (PS), which previously had no reported aptamer [57]. They have since continued their investigation on the PS binding aptamer with more computational and experimental detail with a view to translation for practical use [58].

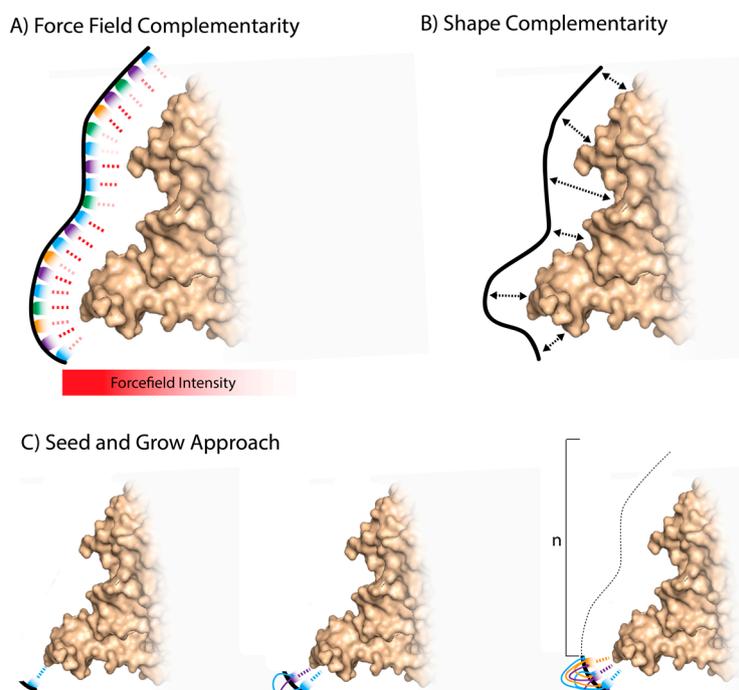


Figure 3. Simplified visualization of the molecular dynamics used in aptamer selections. (A) Force fields, the algorithmically calculated potential energies between atomic components of a nucleotide sequence and a target, are algorithmically calculated and represented here as a dashed red line. The differential opacity represents the variable favorability of each interaction, which is later translated into a docking score; (B) shape complementarity, a simplified representation of the nucleotide sequence interacting with a target based on spatial orientation; (C) the seed and grow approach, a fragment-based method. A single nucleotide is matched with a target and single nucleotides are added in sequence. A probability distribution is used to measure the interactions between the target and linked nucleotides. The target in this figure is derived from crystal structure PDB 3ZH2.

4. Patterning of Libraries

In a SELEX experiment, typically nanomoles of aptamer library or approximately 10^{15} molecules are used. The typical length of a nucleic acid in a library is around 40 bases (total sequence space 4^{40}), so less than one in 1.2 billionth of the sequence space is covered. Aptamers generally require secondary structure to bind their targets, therefore increasing the occurrence of secondary structure in the library should enhance the success rate when selecting for an aptamer. Here, we will outline several approaches that have been taken to pattern aptamer libraries with secondary structures.

RNA aptamer and ribozyme selection analysis has shown that the presence of distinctive secondary structures, such as a stem-loop, enhances the binding affinity [32,59,60]. Such analysis began by observing the nucleotide distributions. Schultes et al. [61] found that functional RNAs have a tendency to have more purine than pyrimidine. This correlation was studied by functional class and phylogenetic domain. It was found that the G+A and G+U content in archaea, bacteria, and eukaryote functional single-stranded RNA showed a similar positive bias and that the bias was inversely proportional to the sequence length [61]. Knight et al. performed a comparative analysis of distantly related and unrelated sequences using simplex to study all possible composition vectors

(G+A, G+U and G+C) of isoleucine aptamer and hammerhead ribozyme [62]. A library size of 6.23×10^9 sequences containing 25% U, 15% C, 20% A, and 40% G could maximize the probability of identifying both motifs (99%). This study demonstrated that adjustment of base composition could be used to lower the total number of candidates in a SELEX (Systematic Evolution of Ligands by Exponential Enrichment) experiment.

Computational methods use sequence information to pattern initial libraries, which results in the evolution of more complex structures. For DNA aptamer selection, Ruff et al. patterned their initial library pool with alternating purine and pyrimidine, which was found to increase the formation of stem-loop structures that bind to streptavidin, immunoglobulin E (IgE), and vascular endothelial growth factor (VEGF) [63]. By sequencing pools from each selection round, they compared the selection efficiency for both random and patterned libraries. The patterned library was significantly enriched relative to the unpatterned library at the 10th round. For IgE, after adding restriction endonuclease to digest the tagged sequences, it was found that the enrichment was further significantly increased. The use of a patterned library in SELEX was able to select specific binders for all three molecules with affinity at nanomolar levels better than those selected from random libraries (streptavidin: $K_D = 105$ nM, IgE: $K_D = 26$ nM, VEGF: $K_D = 45$ nM). These results showed that the use of a patterned library could increase the proportion of active aptamer, speed of selection, and affinity of the resultant aptamers [63].

This alternating purine and pyrimidine patterning strategy was enhanced when Martin et al. used a novel computational method to increase the structural complexity of a DNA library (Table 1) [64]. This patterning method reduced the size of the library, allowing the integration of the entire library onto a microarray, for the identification of a thrombin binding aptamer. The initial library was designed with UNAFold software using three major constraints. First, the first nucleotide of the aptamer must pair with the final one. Second, the number of unpaired bases must fluctuate between 10 and 30 for the 50-nucleotide strands. Third, there must be at least two stretches of unpaired nucleotides. This limited the total number of candidates in the library to 50,000 sequences. The selection results showed that the first six to eight bases of the top 15 sequences resembled thrombin binding aptamer and binding was also specific to thrombin. This demonstrated the effectiveness of using a patterned library on a microarray to select for aptamers.

Table 1. Library design used by Martin et al. [64].

Pattern	Library Design
1	(RY) ₃ -N ₄ -(RY) ₄ -N ₃ -(RY) ₄ -N ₄ -(RY) ₄ -N ₃ -(RY) ₃
2	(RRYY) ₂ -N ₄ -(RRYY)-N ₃ -(RRYY)-N ₄ -(RRYY)-N ₃ -(RRYY)-N ₄ -(RRYY) ₂
3	(RRYY) ₂ -N ₄ -(RRYY) ₃ -N ₄ -(RRYY) ₃ -N ₄ -(RRYY) ₃ -N ₄ -(RRYY) ₂
4	(RRYY) ₂ -N ₄ -(RY) ₃ -N ₄ -(RY) ₃ -N ₄ -(RY) ₃ -N ₄ -(RRYY) ₂

¹ Library designs of different patterns of alternating purine and pyrimidine. Pattern 1 library theoretically has 1.8×10^{19} sequences and Pattern 2 has 3×10^{20} . Pattern 3 has three consecutive purines or pyrimidines, which may allow the formation of quadruplex, while Pattern 4 only allows alternating purines and pyrimidines. R is purine, Y is pyrimidine, and N is a random mixture of purine and pyrimidine. Table adapted from [64].

The 1963 discovery of Hoogsteen base-pairing explains the formation of triplex and quadruplex structures [65,66]. The G-quadruplex structure now holds significant interest due to applications in therapeutics [67,68] and diagnosis [69,70]. A G-quadruplex usually consists of four guanine tracks and a few tetrads [71]. The structural features include a broad surface of π -orbitals above and below the quadruplex that allow hydrophobic binding to targets such as nucleolin [72], hemin [73,74], and light-up fluorogens [75–77].

The binding capacity of G-quadruplex structures was exploited by McManus and Li, who integrated patterning into DNA libraries to select aptamers with quadruplex structures. They patterned the library with the following methods: inclusion of four G-tracts in the library while leaving the rest to be random; simplifying the complicated three-layer G-quadruplex into

a two-layered structure; and four G₂ tracts were designed with three domains of random sequences for loop formation [71,78]. The authors first investigated the effective loop length for the folding of a quadruplex by adding three to seven thymidines between the G-tracts and characterizing by circular dichroism (CD). The CD characteristics of different quadruplex configurations are shown in Table 2. When compared to a completely random library, the G₂ tract library showed peaks at 265, 280, and 295 nm, indicative of G-quadruplex structures (Figure 4), whilst the random library only showed a single peak at 280 nm. The authors also investigated the melting temperatures of individual libraries with different loop lengths at 295 nm. Libraries with a loop length of 3–6 were suitable for the selection of a single-stranded aptamer as their melting temperature did not change at increased concentrations. However, the melting temperature of a library with a loop length of seven increased with concentration. This indicated that it formed multimolecular quadruplexes because the interaction between large loops of different strands has a more regular structure and is not ideal for SELEX, requiring certain structural flexibility in the library. This pioneering work shows how optimizing DNA library parameters can maximize the possibility of selecting active binders.

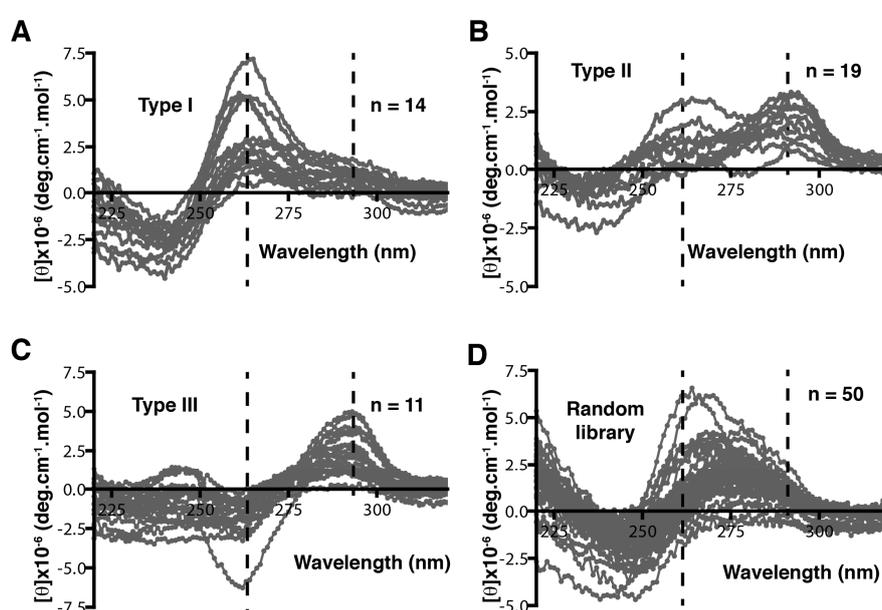


Figure 4. Circular dichroism (CD) characteristics of different quadruplexes and random library. (A) Parallel quadruplex shows positive peaks at 265 and 295 nm; (B) antiparallel quadruplex with glycosidic bond angles of the same orientation shows bimodal spectra as positive peaks at 265 nm and 295 nm of the same intensity; (C) antiparallel quadruplex with glycosidic bond angles of opposite orientation—the CD shows a negative peak at 265 nm and a positive peak at 295 nm; (D) random library shows a peak at 280 nm but no peaks at 265 nm and 295 nm, indicating the absence of a quadruplex. Figure adapted from McManus and Li [78].

Table 2. Major open-source programs for SELEX (Systematic Evolution of Ligands by Exponential Enrichment) HTS (high-throughput sequencing) data analysis.

Program	Operation System	Language	Clustering Method	Validation Experiment
FASTAptamer	Mac/Linux	Perl	Levenshtein distance	HIV-1 Reverse Transcriptase
AptaCluster/AptaGUI	Mac/Linux/PC	Java	LSH and k-mer counting	IL-10RA
APTANI	Linux	Python	Structure motif-based clustering	Murine IL4Ra
AptaTrace	Mac/Linux/PC	C++, Java	Structure motif-based clustering	C-C chemokine receptor
PATTERNITY-seq	No details	No details	Levenshtein distance	Annexin-A2

In a SELEX experiment, the random library provides low sequence space coverage and low structure space representation. Using stem-loop structures and/or patterned libraries can increase sequence space coverage and improve space representation. Although it is difficult to pinpoint the best method to pattern a library, it is clear that patterned libraries can enhance aptamer selection by reducing the time taken to select an aptamer, increasing the success rate of a selection, and improving the binding affinity of isolated aptamers [63].

5. In Silico Aptamer Identification from High-Throughput Sequencing (HTS) Data

5.1. The Trend of Using HTS for Improving SELEX

Initially developed for the purpose of tackling the increasing complexity of whole genome sequencing, HTS technologies have continued to evolve and change the landscape in many fields of biomedical research over the last 10 years [79]. Since the development of the first commercialized 454 sequencer, companies such as Illumina, Ion Torrent, and Oxford Nanopore technologies are all increasing sequencer capacity and reducing cost [80]. In the area of aptamer research, deep, high-throughput, and in-parallel DNA sequencing technologies allow the analysis of millions of sequences found in each round of aptamer selection, and thus open a new avenue for identification and optimization of aptamers [81]. HTS data obtained from each round of the selection can not only be used to monitor the dynamic sequence change of aptamer selection to identify the best-performing sequences in early rounds [82–84], but also as a tool for aptamer scientists to further investigate the enrichment principles of SELEX process such as selection efficiency [85], aptamer–target interactions [86], and mutation landscape [16].

Some of the earliest works applying HTS for identification of aptamers were performed by Schroeder's group in 2010. In the study, they applied a genomic RNA library—overlapping sequences constructed via PCR from the *E. coli* genome—to select against an RNA binding protein named Hfq. 454 sequencing was used to obtain sequence data for two of the last round libraries from the genomic selection and, for comparison purposes, the rounds of another selection that omitted the target binding step, to monitor the amplification variants of the genomic SELEX. By analyzing the HTS data, they successfully identified genomic RNA aptamers and discovered that these aptamers are predominant in the antisense transcripts [87,88]. In the same year, Soh's group pioneered DNA aptamer quantitative selection by applying microfluidic and HTS technologies. They performed three rounds of microfluidic device-assisted selection against platelet-derived growth factor BB (PDGF-BB) and sequenced each round of the selection via high-throughput sequencing. More than 1.7×10^7 sequences from each round of selection were obtained and the enrichment trajectory across different rounds was tracked by analyzing the HTS data. Comparing the sequences obtained from different rounds, they discovered the sequence with the highest affinity did not have the highest copy number in the last round [89]. Schultze et al. [90] confirmed this finding when they discovered that the library convergence in SELEX led to high-performance sequences being outcompeted by weaker-performing sequences that amplify more efficiently during PCR. The best binders tend to enrich rapidly in the very early rounds of selection [90]. Spiga et al. [91] performed HTS and SPR to monitor the binding affinity change and aptamer enrichment for tobramycin selection. They also discovered the most enriched and best binding sequences are visible even after two selection rounds [91]. As the cost of HTS continues to decrease [92], more researchers use it for characterizing multiple selection round libraries to ensure the quality of selected candidates [84,93]. Using the HTS dynamic monitoring method, researchers successfully identified high binding aptamers both for proteins [94] and small molecule targets [91].

5.2. Benchmark Toolkit for HTS SELEX Analysis

Besides the cost, one of the major hindrances in early years to generalize HTS methods for aptamer identification was the difficulty of processing large amounts of sequence data. However, multiple open-source/paid bioinformatics tools have been developed specifically for aptamer scientists.

The initial step in processing HTS data from a sequenced SELEX pool is to remove the adapter, barcode, or constant region from the sequences. After this pre-processing, a tool to count the sequence frequency is required. Previously, aptamer research groups used genomic informatics software packages such as Tallymer [95,96] or RazerS [90,97], or designed in-house programs to fulfill the counting requirement [89,98].

Galaxy Project is a platform that provides fundamental bioinformatic tools for bench scientists who may not have a background in bioinformatics. Thiel et al. recently developed workflows based on this Galaxy Project for handling HTS SELEX data to perform pre-processing steps [99,100]. This tool also allows researchers to remove adapter/barcode/primer regions from sequences; identify and remove sequences with mismatches within the primer region; set a variable region length cutoff; and count the number of duplicate reads. Another benefit of the Galaxy workflow is it is “ready to use” and “easy to access” for an open-source, web-based platform. However, the Galaxy web service does not currently provide analysis for motif-based clustering as the platform was designed for general genomic projects.

An easy-to-use, aptamer-specific bioinformatics tool to address the clustering based on primary sequence is FASTAptamer [101]. FASTAptamer consists of a library of modular Perl scripts and is compatible across UNIX-like systems (or a Windows system with a Perl interpreter installed). Count, compare, cluster, enrich, and search are the five major script modules. By using these modules, users can count, normalize, and rank the sequence reads in a FASTQ file and group these sequences into families based on Levenshtein distance, as well as determine the SELEX enrichment across multiple selection rounds [101]. PATTERNITY-seq, developed by Ducongé’s group [102], is another package that uses sequence pattern clustering based on Levenshtein distance. They validated this approach by re-analyzing the data from a previously published cell-SELEX against Annexin-A2 [103].

AptaCluster [104] is similar to FASTAptamer but based on the local sensitive hashing (LSH) method, which is capable of comparing sequences with a reduced number of dimensions (Figure 5). Iterative rounds of comparison within groups of aptamers are used to cluster aptamer sequences. By using such a method, the computational time required for AptaCluster is less than FASTAptamer. However, this method cannot be applied as a sequence pool containing various sizes. FASTAptamer and AptaCluster are purely text-based tools, whereas a program called AptaGUI that can be used alongside AptaCluster includes a graphical user interface (GUI) for the dynamic visualization of HTS SELEX data [105]. An alternative to AptaGUI is the paid platform COMPAS, developed by AptaIT GmbH. This also contains GUI for the navigation of the HTS data, but many of the operational details are proprietary and the company program is only available in conjunction with the purchase of their selection service.

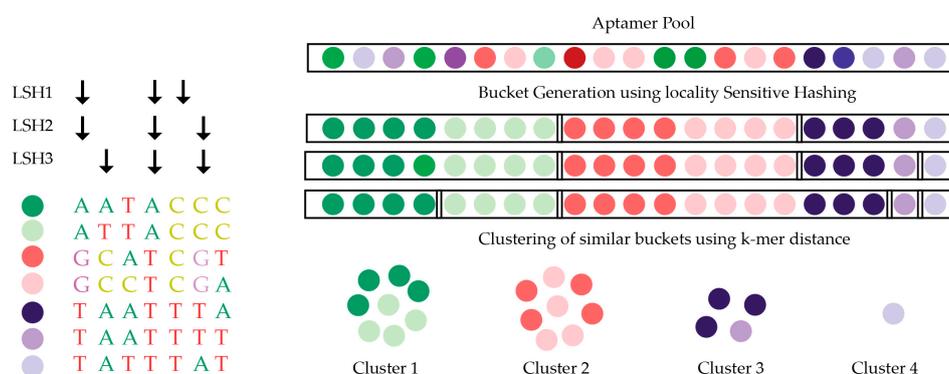


Figure 5. Illustration of the AptaCluster algorithm. Each colored sphere represents an individual sequence in the library and the similar colors represent related sequence. AptaCluster clusters the library pool into different sets of similar sequences based on locality sensitive hashing (LSH). The black arrows represent the user-defined number of nucleotide positions, which are sampled to generate input for the hash function. Figure adapted from Honika [16].

5.3. Structure Motif Clustering-Based Tools

Most of the previously mentioned bioinformatics tools do not include functions for prediction and clustering of HTS data for structure-based methods. Furthermore, they do not allow elucidation of complex motifs and important pre-processing steps for initial analysis of the data generated during SELEX.

Structure prediction programs such as Mfold [106] have been used to analyze low-throughput sequence data. Mfold predicts the secondary structure of single-stranded nucleic acids by energy minimization. Even though “bulk” servers of Mfold can analyze hundreds of sequences at once, it is, however, difficult to handle structure prediction on the HTS scale.

A recently developed platform for structure motif clustering is AptaTrace [107]. Based on the secondary structure prediction from SFOLD [108,109], AptaTrace applies this information into all of the sequences input to the program. This allows for the prediction of a specific structure for each k-mer in each selection round and ranking by predicted significant structural enrichment. APTANI is a similar program, able to cluster sequence motifs based on secondary structure prediction. It uses RNAsubopt from the Vienna RNA package [110] and predicts using sub-structures, apical loops, bulge loops, and intra-strand loops. This method was validated using a SELEX against IL4Ra. Using APTANI, an aptamer was identified in one round, which previously required five rounds.

The speed of HTS technology adoption has motivated the development of particular tools to assist HTS-based SELEX and identify better aptamer candidates (Table 2). Even though many approaches still lack multiple validations, using HTS to replace conventional sequencing methods for aptamer development is the trend. Recent progress in this field shows the potential for developing an all-in-one bioinformatics tools for aptamer researchers.

6. In Silico Aptamer Optimization

Aptamers have been isolated with both high affinity and high specificity for binding to their selected targets [5,111]. SELEX is an efficient method of isolating aptamers; however, following selection an aptamer scientist must always ask “Have I isolated the best possible aptamer sequence?” The library used for SELEX generally has a random region of around 40 bases [4], and typically only a few nanomoles can be used for the initial selection round. This represents a sequence space coverage of one in 80 billion. From this incredibly small sequence space coverage, it is unlikely that one will select the single best aptamer sequence. Sequence adaptation via mutation may account for some sequence space searching; however, as selection for SELEX is relatively low resolution [112], it is difficult to resolve the very best aptamer sequence.

Bioinformatic approaches have been used to improve the affinity of aptamers. As highlighted earlier, due to low selection pressure classical SELEX is unlikely to resolve the very best aptamer sequences. Therefore, each individual aptamer generated using a bioinformatics approach must be individually assayed for binding affinity, which can be labor-intensive and time-consuming. DNA microarrays consist of many features or spots on a glass slide, each feature containing many copies of a unique DNA sequence. This high-throughput technology allows for simultaneous assay of many aptamer sequences via incubation with fluorescent target.

In 2007, Katilius et al. used DNA microarrays to optimize and explore the surrounding sequence space of an aptamer against immunoglobulin E (IgE) [113]. Variations of the aptamer sequence with single, double and a selection of triple point mutations were synthesized onto a DNA microarray and assayed with Alexa Fluor 647 labeled IgE. This mutational analysis highlighted the conserved and unconserved base positions in the aptamer sequence. One aptamer variant showed mild affinity improvement ($K_D = 4.1$ nM) when compared to the original aptamer sequence ($K_D = 4.7$ nM) [113].

Platt et al. analyzed the sequence activity relationship of a set of G-quadruplex thrombin binding aptamers using DNA microarray technology [114]. The combinatorial landscape was probed via two methods. The first method investigated two internal loops of the G-quadruplex with 2-3 base random regions (GGGGAGTAGG(X₂₋₃)GGTGTTGG(X₂₋₃)GGGGCTCCCC, where X denotes the bases varied).

The second method investigated the hairpin in which the G-quadruplex is nested within a section using pseudo-random variants $((X_8)GGTT(X_{2-4})GGTTGGGG(X_6))$, where X denotes the bases varied). Despite this search through sequence space, the tightest binding novel aptamer ($K_D = 28$ nM) had a lower affinity than the original ThB aptamer ($K_D = 26$ nM) [114].

Knight et al. combined a DNA aptamer microarray assay with *in silico* closed-loop aptameric directed evolution (CLADE) to select for aptamers against the natively fluorescent target allophycocyanin (APC) [115]. Five hundred control aptamer and 5500 test pool aptamers of 30 nucleotides were synthesized onto a DNA microarray for each round. The initial test pool for the first round was randomly generated. The test pool aptamers were assayed on the microarray for APC binding and ranked according to binding score. The top four aptamers were then subjected to point mutations and insertion–deletion events to give rise to a new 5500 test pool for synthesis onto a DNA microarray and use in the next round of selection. Nine rounds of CLADE were performed and the resulting aptamers characterized and phylogenetically analyzed. The CLADE strategy was successful with the tightest binding aptamer had a SPR determined K_D value of around 2 nM [115]. Although high-affinity aptamers were isolated, the cost of nine microarrays would be much greater than the cost of an average SELEX experiment.

Expanding upon this work, Rowe et al. used the CLADE approach and tested the three diversification systems: mutation, recombination, and statistical binding prediction [116]. Over five CLADE selection rounds, aptamers were evolved to bind to glucose-6-phosphate dehydrogenase. The tightest binding aptamer was isolated using the recombination diversification system and had a K_D of 245 nM.

In 2012, Nonaka et al. used an *in silico* system to improve the affinity of the VEap121 aptamer against VEGF [117]. Interestingly, this study did not use DNA microarrays but instead used SPR to assay every individual aptamer. Three rounds of improvement were performed. Each round consisted of adaptation, SPR determination of K_D value, and selection of the five tightest binding aptamers to seed the next round. For the first generation, 10 mutants of VEap121 were generated, each with several mutations, where the guanine bases were conserved to retain the G-quadruplex structure. For the second generation, the five tightest binding aptamers from G1 as determined by SPR were replicated relative to their binding affinity to yield 20 sequences. These were then randomly paired to undergo single-point crossover and two single-base mutations, randomly introduced. For the third generation, the five tightest binding aptamers from both G1 and G2 as determined by SPR were crossed with VEap121 at a random point and two single-point mutations were randomly introduced. This process was repeated three times in order to produce the third generation of 20 sequences. This process produced four aptamers with a tighter binding ($K_D = 0.3, 1.5, 1.7$ and 2.4 nM) than the original VEap121 aptamer ($K_D = 4.7$ nM) [117].

In 2016 Kinghorn et al. reported a novel strategy of aptamer affinity maturation by library resampling from SELEX sequence data [118]. This approach relies on the principle that classical SELEX is unlikely to select the best possible aptamer, but is highly likely to select family members of the best possible aptamer. The sequence of the best possible aptamer is hidden within the sequences of its family members. To make use of this aptamer family information, the authors coded the bioinformatic software “Resample”, which uses information from a SELEX experiment in terms of an aptamer family motif and any available folding information. This information is used to generate a novel library that consists of every possible aptamer permutation within the aptamer family. This library is focused on a particular area of sequence space, representing it thoroughly while still having a library size small enough to fit onto a DNA microarray for screening (Figure 6). To demonstrate this process, the sequence data from a previous selection against the malarial antigen *Plasmodium falciparum* lactate dehydrogenase (PfLDH) was input into Resample to generate a library of 186,624 novel aptamer sequences within the specified aptamer family. This library was ordered on a DNA microarray that was incubated with 50 nM Alexa Fluor 555 labeled PfLDH (target) and 1 μ M Alexa Fluor 647 labeled human lactate dehydrogenase B (counter-target) and washed and scanned to measure both binding affinity and binding specificity for all aptamers. The lead candidates were further characterized

using microscale thermophoresis to show an order of magnitude improvement in binding affinity K_D value. The authors provide Resample as a hosted web service (website available: <http://resample.azurewebsites.net>) and state that affinity maturation using Resample round should just take two days, excluding microarray shipping time. While many microarray aptamer optimization processes use mutation and recombination, taking many small evolutionary steps to arrive at an optimum, Resample takes all possible selection sequence information and takes a single evolutionary leap to an optimum. In this way, a single Resample diversification round can be used to hone in on the sequence space containing the best possible aptamer sequence [118].

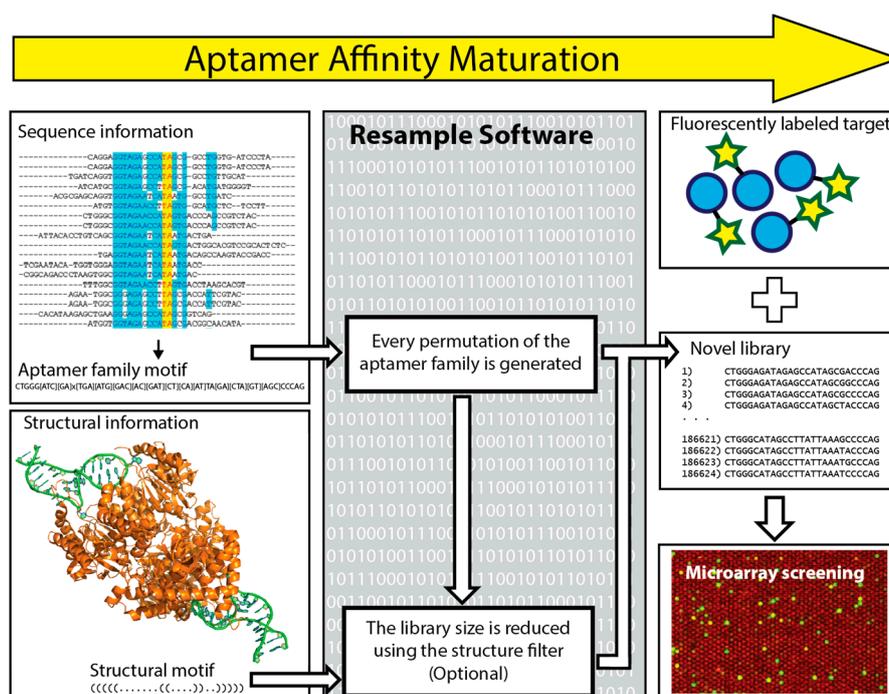


Figure 6. The aptamer affinity maturation process. The yellow arrow designates the aptamer affinity maturation process, in which data from a SELEX experiment is utilized by bioinformatic software to Create a novel aptamer library which undergoes a microarray screening process. Sequence information and, if available, folding information (optional) from a SELEX experiment is used as input for the resample software. Resample outputs a novel focused aptamer library that is small enough to fit on a DNA microarray. The microarray is incubated with a fluorescently labeled target before washing and scanning to measure the binding strength of all aptamers. Figure adapted from Kinghorn et al. [118].

7. Conclusions and Future Perspectives

Simulation of aptamer selection has given insight into the SELEX process, including the optimum protein/target concentration, the detrimental effect of low relative partitioning efficiency, the effect of background binding, and the stochastic nature of SELEX. The area least accurately modeled in SELEX simulations is aptamer binding. Development of more accurate aptamer binding models and applying them in simulation may lead to new insights into SELEX.

In silico approaches have been used to classify molecular interactions between binding macromolecules into discrete categories. Each categorized interaction is assigned a probability and level of importance, which is translated into a score. By grouping all the scores, it is possible to distinguish a poor binding interaction from a strong interaction. Most programs have been developed for protein–protein interactions, but in recent years useful software has been devised to investigate DNA–target interactions. These in silico tools for both whole aptamer and fragment approaches have aided aptamer scientists in improving their selections and identifying novel high-affinity aptamers.

It is commonly known that the huge number of candidates in a random nucleic acid library cannot be covered by a single SELEX experiment. One of the solutions to reduce that number or increase the coverage of selection is to introduce patterns into the library. In silico methods include defining alternating purine and pyrimidine patterns leading to the increase in occurrence of stem-loop structures or more complicated structures such as quadruplexes. The design could also be combined with in vitro experiments to access the structural diversity of certain patterned libraries by CD and NMR. Such approaches will effectively help to increase the success rate of identifying active binders in the selection process.

HTS technology shows high potential to replace the cloning and Sanger sequencing methods applied in traditional SELEX. By integrating an HTS step into SELEX, researchers can successfully reduce the selection rounds and the need for post-selection experiments to identify optimal aptamer sequences. These advantages of HTS technology encourage the rapid development of aptamer-based bioinformatic tools. There are several software packages and databases customized for aptamer scientists (Table 2) to analyze the large amount of HTS data based on different strategies. It will be useful to consider how best to compare these tools, using the same batch of data with multiple validations from different research groups.

In silico aptamer optimization has not been widely adopted. This may be due to the observation that most studies either achieve only mild affinity improvements or that the optimization method, while successful, is prohibitively expensive. Nonaka et al. achieved binding improvement of an order of magnitude by using a low-cost method, albeit labor-intensive in SPR measurements [117]. Kinghorn et al. achieved binding improvement of an order of magnitude with a low-cost method that can be performed in two days, excluding microarray shipping time [118]. Many aptamer optimization studies are stand-alones without follow-up or verification by other research groups. For the aptamer community to adopt in silico aptamer optimization, replicate studies need to be performed to strengthen and further validate in silico aptamer optimization methods.

Bioinformatic approaches have been used to improve both aptamers and their selection. In this review we have outlined a broad range of aptamer bioinformatics techniques including simulation of aptamer selection, aptamer selection by molecular dynamics, patterning of libraries, identification of lead aptamers from HTS data, and in silico aptamer optimization. Aptamers are particularly suited to bioinformatic techniques and their development and use can benefit aptamer scientific community.

Acknowledgments: This work was supported in part by the Hong Kong General Research Fund under grants 17127515 and 17163416 to JAT.

Author Contributions: Andrew B. Kinghorn wrote Section 2. Simulation of Aptamer Selection and Section 6. In Silico Aptamer Optimization as well as manuscript editing. Lewis A. Fraser wrote Section 3. Aptamer Selection by Molecular Dynamics as well as manuscript editing. Shaolin Liang wrote Section 5. In Silico Aptamer Identification from High-Throughput Sequencing (HTS) Data as well as manuscript editing. Simon Chi-Chin Shiu wrote Section 4. Patterning of libraries as well as manuscript editing. Julian A. Tanner edited and directed writing of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ellington, A.D.; Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818–822. [[CrossRef](#)] [[PubMed](#)]
2. Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505–510. [[CrossRef](#)] [[PubMed](#)]
3. Cruz-Toledo, J.; McKeague, M.; Zhang, X.; Giamberardino, A.; McConnell, E.; Francis, T.; DeRosa, M.C.; Dumontier, M. Aptamer base: A collaborative knowledge base to describe aptamers and SELEX experiments. *Database* **2012**. [[CrossRef](#)] [[PubMed](#)]
4. McKeague, M.; McConnell, E.M.; Cruz-Toledo, J.; Bernard, E.D.; Pach, A.; Mastronardi, E.; Zhang, X.; Beking, M.; Francis, T.; Giamberardino, A. Analysis of in vitro aptamer selection parameters. *J. Mol. Evol.* **2015**, *81*, 150–161. [[CrossRef](#)] [[PubMed](#)]

5. Cheung, Y.-W.; Kwok, J.; Law, A.W.; Watt, R.M.; Kotaka, M.; Tanner, J.A. Structural basis for discriminatory recognition of *Plasmodium lactate* dehydrogenase by a DNA aptamer. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 15967–15972. [[CrossRef](#)] [[PubMed](#)]
6. Irvine, D.; Tuerk, C.; Gold, L. SELEXION: Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.* **1991**, *222*, 739–761. [[CrossRef](#)]
7. Vant-Hull, B.; Payano-Baez, A.; Davis, R.H.; Gold, L. The mathematics of SELEX against complex targets. *J. Mol. Biol.* **1998**, *278*, 579–597. [[CrossRef](#)] [[PubMed](#)]
8. Homann, M.; Göringer, H.U. Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucleic Acids Res.* **1999**, *27*, 2006–2014. [[CrossRef](#)] [[PubMed](#)]
9. Coulter, L.R.; Landree, M.A.; Cooper, T.A. Identification of a new class of exonic splicing enhancers by in vivo selection. *Mol. Cell. Biol.* **1997**, *17*, 2143–2150. [[CrossRef](#)] [[PubMed](#)]
10. Chen, C.-K.; Kuo, T.-L.; Chan, P.-C.; Lin, L.-Y. Subtractive SELEX against two heterogeneous target samples: Numerical simulations and analysis. *Comput. Biol. Med.* **2007**, *37*, 750–759. [[CrossRef](#)] [[PubMed](#)]
11. Chen, C.-K. Complex SELEX against target mixture: Stochastic computer model, simulation, and analysis. *Comput. Meth. Prog. Biol.* **2007**, *87*, 189–200. [[CrossRef](#)] [[PubMed](#)]
12. Wang, J.; Rudzinski, J.F.; Gong, Q.; Soh, H.T.; Atzberger, P.J. Influence of target concentration and background binding on in vitro selection of affinity reagents. *PLoS ONE* **2012**, *7*, e43940. [[CrossRef](#)] [[PubMed](#)]
13. Vant-Hull, B.; Gold, L.; Zichi, D.A. Theoretical principles of in vitro selection using combinatorial nucleic acid libraries. *Curr. Protoc. Nucleic Acid Chem.* **2000**. [[CrossRef](#)]
14. Zhao, Y.; Granas, D.; Stormo, G.D. Inferring binding energies from selected binding sites. *PLoS Comput. Biol.* **2009**, *5*, e1000590. [[CrossRef](#)] [[PubMed](#)]
15. Spill, F.; Weinstein, Z.B.; Shemirani, A.I.; Ho, N.; Desai, D.; Zaman, M.H. Controlling uncertainty in aptamer selection. *Proc. Natl. Acad. Sci. USA* **2016**. [[CrossRef](#)] [[PubMed](#)]
16. Hoinka, J.; Berezhnoy, A.; Dao, P.; Sauna, Z.E.; Gilboa, E.; Przytycka, T.M. Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery. *Nucleic Acids Res.* **2015**, *43*, 5699–5707. [[CrossRef](#)] [[PubMed](#)]
17. Oh, I.S.; Lee, Y.-G.; McKay, R. Simulating chemical evolution. In Proceedings of the IEEE Congress on Evolutionary Computation, New Orleans, LA, USA, 5–8 July 2011; pp. 2717–2724. [[CrossRef](#)]
18. Kauffman, S.; Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **1987**, *128*, 11–45. [[CrossRef](#)]
19. Wedge, D.C.; Rowe, W.; Kell, D.B.; Knowles, J. In silico modelling of directed evolution: Implications for experimental design and stepwise evolution. *J. Theor. Biol.* **2009**, *257*, 131–141. [[CrossRef](#)] [[PubMed](#)]
20. Deem, M.W.; Lee, H.Y. Sequence space localization in the immune system response to vaccination and disease. *Phys. Rev. Lett.* **2003**, *91*, 068101. [[CrossRef](#)] [[PubMed](#)]
21. Hall, M.; Christensen, K.; di Collobiano, S.A.; Jensen, H.J. Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution. *Phys. Rev. E* **2002**, *66*, 011904. [[CrossRef](#)] [[PubMed](#)]
22. Kauffman, S.; Macready, W. Technological evolution and adaptive organizations: Ideas from biology may find applications in economics. *Complexity* **1995**, *1*, 26–43. [[CrossRef](#)]
23. Klussmann, S. Functional oligonucleotides and their applications. In *The Aptamer Handbook*; John Wiley & Sons: New York, NY, USA, 2006.
24. Kauffman, S.A. *The Origins of Order: Self Organization and Selection in Evolution*; Oxford University Press: Oxford, UK, 1992; pp. 61–100.
25. Altenberg, L. Evolving better representations through selective genome growth. In Proceedings of the First IEEE Conference on IEEE World Congress on Computational Intelligence, Evolutionary Computation, Orlando, FL, USA, 27–29 June 1994; pp. 182–187.
26. Kinghorn, A.B.; Tanner, J.A. Selective Phenome Growth Adapted Model: A Novel Landscape to Represent Aptamer Ligand Binding. *Complexity* **2017**, *2017*, 1–12. [[CrossRef](#)]
27. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evolut. Comput.* **1997**, *1*, 67–82. [[CrossRef](#)]
28. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L., Jr.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616. [[CrossRef](#)] [[PubMed](#)]

29. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71.
30. MacKerell, A.D.; Nilsson, L. Molecular dynamics simulations of nucleic acid–protein complexes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194–199. [[CrossRef](#)] [[PubMed](#)]
31. Gurtovenko, A.A.; Vattulainen, I. Pore formation coupled to ion transport through lipid membranes as induced by transmembrane ionic charge imbalance: Atomistic molecular dynamics study. *J. Am. Chem. Soc.* **2005**, *127*, 17570–17571. [[CrossRef](#)] [[PubMed](#)]
32. Carothers, J.M.; Oestreich, S.C.; Szostak, J.W. Aptamers selected for higher-affinity binding are not more specific for the target ligand. *J. Am. Chem. Soc.* **2006**, *128*, 7929–7937. [[CrossRef](#)] [[PubMed](#)]
33. Gevertz, J.; Gan, H.H.; Schlick, T. In vitro RNA random pools are not structurally diverse: A computational analysis. *RNA* **2005**, *11*, 853–863. [[CrossRef](#)]
34. Chushak, Y.; Stone, M.O. In silico selection of RNA aptamers. *Nucleic Acids Res.* **2009**, *37*, e87. [[CrossRef](#)] [[PubMed](#)]
35. Hua, W.P.; Linb, H.T.; Tsaia, J.J.; Chenc, W.Y. Investigating interactions between proteins and nucleic acids by computational approaches. In *Computational Methods with Applications in Bioinformatics Analysis*; World Scientific: Singapore, 2017; p. 98.
36. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195–201. [[CrossRef](#)] [[PubMed](#)]
37. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)] [[PubMed](#)]
38. Van Dijk, M.; Bonvin, A.M. 3D-DART: A DNA structure modelling server. *Nucleic Acids Res.* **2009**, *37*, W235–W239. [[CrossRef](#)] [[PubMed](#)]
39. Popena, M.; Szachniuk, M.; Antczak, M.; Purzycka, K.J.; Lukasiak, P.; Bartol, N.; Blazewicz, J.; Adamiak, R.W. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **2012**, *40*, e112. [[CrossRef](#)] [[PubMed](#)]
40. Setny, P.; Bahadur, R.P.; Zacharias, M. Protein–DNA docking with a coarse-grained force field. *BMC Bioinform.* **2012**, *13*, 228. [[CrossRef](#)] [[PubMed](#)]
41. Das, R.; Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14664–14669. [[CrossRef](#)] [[PubMed](#)]
42. Tsui, V.; Case, D.A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275–291. [[CrossRef](#)]
43. Zhang, S.; Kumar, K.; Jiang, X.; Wallqvist, A.; Reifman, J. DOVIS: An implementation for high-throughput virtual screening using AutoDock. *BMC Bioinform.* **2008**, *9*, 126. [[CrossRef](#)] [[PubMed](#)]
44. Chen, R.; Li, L.; Weng, Z. ZDOCK: An initial-stage protein-docking algorithm. *Proteins* **2003**, *52*, 80–87. [[CrossRef](#)] [[PubMed](#)]
45. Kumar, J.V.; Chen, W.-Y.; Tsai, J.J.; Hu, W.-P. Molecular simulation methods for selecting thrombin-binding aptamers. In *Information Technology Convergence*; Springer: Dordrecht, The Netherlands, 2013; Volume 253, pp. 743–749.
46. White, R.R.; Shan, S.; Rusconi, C.P.; Shetty, G.; Dewhirst, M.W.; Kontos, C.D.; Sullenger, B.A. Inhibition of rat corneal angiogenesis by a nuclease-resistant RNA aptamer specific for angiopoietin-2. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5028–5033. [[CrossRef](#)] [[PubMed](#)]
47. Sarraf-Yazdi, S.; Mi, J.; Moeller, B.J.; Niu, X.; White, R.R.; Kontos, C.D.; Sullenger, B.A.; Dewhirst, M.W.; Clary, B.M. Inhibition of in vivo tumor angiogenesis and growth via systemic delivery of an angiopoietin 2-specific RNA aptamer. *J. Surg. Res.* **2008**, *146*, 16–23. [[CrossRef](#)] [[PubMed](#)]
48. Hu, W.-P.; Kumar, J.V.; Huang, C.-J.; Chen, W.-Y. Computational selection of RNA aptamer against angiopoietin-2 and experimental evaluation. *BioMed Res. Int.* **2015**, *2015*, 658712. [[CrossRef](#)] [[PubMed](#)]
49. Maisonpierre, P.C.; Suri, C.; Jones, P.F.; Bartunkova, S.; Wiegand, S.J.; Radziejewski, C.; Compton, D.; McClain, J.; Aldrich, T.H.; Papadopoulos, N. Angiopoietin-2, a natural antagonist for Tie2 that disrupts in vivo angiogenesis. *Science* **1997**, *277*, 55–60. [[CrossRef](#)] [[PubMed](#)]
50. Holash, J.; Maisonpierre, P.; Compton, D.; Boland, P.; Alexander, C.; Zagzag, D.; Yancopoulos, G.; Wiegand, S. Vessel cooption, regression, and growth in tumors mediated by angiopoietins and VEGF. *Science* **1999**, *284*, 1994–1998. [[CrossRef](#)] [[PubMed](#)]
51. Lin, H.; Zhang, W.; Jia, S.; Guan, Z.; Yang, C.J.; Zhu, Z. Microfluidic approaches to rapid and efficient aptamer selection. *Biomicrofluid* **2014**, *8*, 041501. [[CrossRef](#)] [[PubMed](#)]

52. Shangguan, D.; Bing, T.; Zhang, N. *Cell-SELEX: Aptamer selection against whole cells, in Aptamers Selected by Cell-SELEX for Theranostics*; Springer: Heidelberg, Germany, 2015; pp. 13–33.
53. Mallikaratchy, P. Evolution of Complex Target SELEX to Identify Aptamers against Mammalian Cell-Surface Antigens. *Molecules* **2017**, *22*, 215. [[CrossRef](#)] [[PubMed](#)]
54. Rabal, O.; Pastor, F.; Villanueva, H.; Soldevilla, M.M.; Hervás-Stubbs, S.; Oyarzabal, J. In Silico Aptamer Docking Studies: From a Retrospective Validation to a Prospective Case Study/TIM3 Aptamers Binding. *Mol. Ther. Nucleic Acids* **2016**, *5*, e376. [[CrossRef](#)] [[PubMed](#)]
55. Huang, Y.; Liu, S.; Guo, D.; Li, L.; Xiao, Y. A novel protocol for three-dimensional structure prediction of RNA-protein complexes. *Sci. Rep.* **2013**, *3*, 1887. [[CrossRef](#)] [[PubMed](#)]
56. Berendsen, H.J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56. [[CrossRef](#)]
57. Tseng, C.Y.; Ashrafuzzaman, M.; Mane, J.Y.; Kaptj, J.; Mercer, J.R.; Tuszynski, J.A. Entropic Fragment-Based Approach to Aptamer Design. *Chem. Biol. Drug Des.* **2011**, *78*, 1–13. [[CrossRef](#)] [[PubMed](#)]
58. Ashrafuzzaman, M.; Tseng, C.-Y.; Kaptj, J.; Mercer, J.R.; Tuszynski, J.A. A computationally designed DNA aptamer template with specific binding to phosphatidylserine. *Nucleic Acid Ther.* **2013**, *23*, 418–426. [[CrossRef](#)]
59. Carothers, J.M.; Oestreich, S.C.; Davis, J.H.; Szostak, J.W. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.* **2004**, *126*, 5130–5137. [[CrossRef](#)] [[PubMed](#)]
60. Davis, J.H.; Szostak, J.W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11616–11621. [[CrossRef](#)] [[PubMed](#)]
61. Schultes, E.; Hrabec, P.T.; LaBean, T.H. Global similarities in nucleotide base composition among disparate functional classes of single-stranded RNA imply adaptive evolutionary convergence. *RNA* **1997**, *3*, 792–806. [[PubMed](#)]
62. Knight, R.; De Sterck, H.; Markel, R.; Smit, S.; Oshmyansky, A.; Yarus, M. Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids. *Nucleic Acids Res.* **2005**, *33*, 5924–5935. [[CrossRef](#)] [[PubMed](#)]
63. Ruff, K.M.; Snyder, T.M.; Liu, D.R. Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure. *J. Am. Chem. Soc.* **2010**, *132*, 9453–9464. [[CrossRef](#)] [[PubMed](#)]
64. Martin, J.A.; Mirau, P.A.; Chushak, Y.; Chavez, J.L.; Naik, R.R.; Hagen, J.A.; Kelley-Loughnane, N. Single-Round Patterned DNA Library Microarray Aptamer Lead Identification. *J. Anal. Methods Chem.* **2015**, *2015*, 137489. [[CrossRef](#)] [[PubMed](#)]
65. Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **1963**, *16*, 907–916. [[CrossRef](#)]
66. Frank-Kamenetskii, M.D.; Mirkin, S.M. Triplex DNA structures. *Annu. Rev. Biochem.* **1995**, *64*, 65–95. [[CrossRef](#)] [[PubMed](#)]
67. De Nicola, B.; Lech, C.J.; Heddi, B.; Regmi, S.; Frasson, I.; Perrone, R.; Richter, S.N.; Phan, A.T. Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res.* **2016**, *44*, 6442–6451. [[CrossRef](#)] [[PubMed](#)]
68. Corey, D.R. Telomeres and telomerase: From discovery to clinical trials. *Chem. Biol.* **2009**, *16*, 1219–1223. [[CrossRef](#)] [[PubMed](#)]
69. Bourdoncle, A.; Estévez Torres, A.; Gosse, C.; Lacroix, L.; Vekhoff, P.; Le Saux, T.; Jullien, L.; Mergny, J.L. Quadruplex-Based Molecular Beacons as Tunable DNA Probes. *J. Am. Chem. Soc.* **2006**, *128*, 11094–11105. [[CrossRef](#)] [[PubMed](#)]
70. Zhu, J.; Zhang, L.; Li, T.; Dong, S.; Wang, E. Enzyme-free unlabeled DNA logic circuits based on toehold-mediated strand displacement and split G-quadruplex enhanced fluorescence. *Adv. Mater.* **2013**, *25*, 2440–2444. [[CrossRef](#)] [[PubMed](#)]
71. Webba da Silva, M. Geometric formalism for DNA quadruplex folding. *Chemistry* **2007**, *13*, 9738–9745. [[CrossRef](#)] [[PubMed](#)]
72. Bates, P.J.; Kahlon, J.B.; Thomas, S.D.; Trent, J.O.; Miller, D.M. Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. *J. Biol. Chem.* **1999**, *274*, 26369–26377. [[CrossRef](#)] [[PubMed](#)]
73. Travascio, P.; Li, Y.; Sen, D. DNA-enhanced peroxidase activity of a DNA-aptamer-hemin complex. *Chem. Biol.* **1998**, *5*, 505–517. [[CrossRef](#)]

74. Sen, D.; Poon, L.C. RNA and DNA complexes with hemin [Fe(III) heme] are efficient peroxidases and peroxxygenases: How do they do it and what does it mean? *Crit. Rev. Biochem. Mol. Biol.* **2011**, *46*, 478–492. [[CrossRef](#)] [[PubMed](#)]
75. Baugh, C.; Grate, D.; Wilson, C. 2.8 Å crystal structure of the malachite green aptamer11Edited by J. A. Doudna. *J. Mol. Biol.* **2000**, *301*, 117–128. [[CrossRef](#)] [[PubMed](#)]
76. Warner, K.D.; Chen, M.C.; Song, W.; Strack, R.L.; Thorn, A.; Jaffrey, S.R.; Ferre-D'Amare, A.R. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* **2014**, *21*, 658–663. [[CrossRef](#)] [[PubMed](#)]
77. Trachman, R.J., 3rd; Demeshkina, N.A.; Lau, M.W.L.; Panchapakesan, S.S.S.; Jeng, S.C.Y.; Unrau, P.J.; Ferre-D'Amare, A.R. Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nat. Chem. Biol.* **2017**, *13*, 807–813. [[CrossRef](#)] [[PubMed](#)]
78. McManus, S.A.; Li, Y. Assessing the amount of quadruplex structures present within G(2)-tract synthetic random-sequence DNA libraries. *PLoS ONE* **2013**, *8*, e64131. [[CrossRef](#)] [[PubMed](#)]
79. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [[CrossRef](#)] [[PubMed](#)]
80. Marian, A.J. Sequencing your genome: What does it mean? *Methodist Debaquey Cardiovasc. J.* **2014**, *10*, 3–6. [[CrossRef](#)] [[PubMed](#)]
81. Blind, M.; Blank, M. Aptamer selection technology and recent advances. *Mol. Ther. Nucleic Acids* **2015**, *4*, e223. [[CrossRef](#)] [[PubMed](#)]
82. Berezhnoy, A.; Stewart, C.A.; McNamara, J.O.; Thiel, W.; Giangrande, P.; Trinchieri, G.; Gilboa, E. Isolation and optimization of murine IL-10 receptor blocking oligonucleotide aptamers using high-throughput sequencing. *Mol. Ther.* **2012**, *20*, 1242–1250. [[CrossRef](#)] [[PubMed](#)]
83. Thiel, W.H.; Bair, T.; Peek, A.S.; Liu, X.; Dassie, J.; Stockdale, K.R.; Behlke, M.A.; Miller, F.J., Jr.; Giangrande, P.H. Rapid identification of cell-specific, internalizing RNA aptamers with bioinformatics analyses of a cell-based aptamer selection. *PLoS ONE* **2012**, *7*, e43836. [[CrossRef](#)] [[PubMed](#)]
84. Valenzano, S.; De Girolamo, A.; DeRosa, M.C.; McKeague, M.; Schena, R.; Catucci, L.; Pascale, M. Screening and Identification of DNA Aptamers to Tyramine Using in Vitro Selection and High-Throughput Sequencing. *ACS Comb. Sci.* **2016**, *18*, 302–313. [[CrossRef](#)] [[PubMed](#)]
85. Takahashi, M.; Wu, X.; Ho, M.; Chomchan, P.; Rossi, J.J.; Burnett, J.C.; Zhou, J. High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency. *Sci. Rep.* **2016**, *6*, 33697. [[CrossRef](#)] [[PubMed](#)]
86. Dupont, D.M.; Larsen, N.; Jensen, J.K.; Andreasen, P.A.; Kjems, J. Characterisation of aptamer-target interactions by branched selection and high-throughput sequencing of SELEX pools. *Nucleic Acids Res.* **2015**, *43*, e139. [[CrossRef](#)] [[PubMed](#)]
87. Zimmermann, B.; Gesell, T.; Chen, D.; Lorenz, C.; Schroeder, R. Monitoring genomic sequences during SELEX using high-throughput sequencing: Neutral SELEX. *PLoS ONE* **2010**, *5*, e9169. [[CrossRef](#)] [[PubMed](#)]
88. Lorenz, C.; Gesell, T.; Zimmermann, B.; Schoeberl, U.; Bilusic, I.; Rajkowitsch, L.; Waldsich, C.; von Haeseler, A.; Schroeder, R. Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res.* **2010**, *38*, 3794–3808. [[CrossRef](#)] [[PubMed](#)]
89. Cho, M.; Xiao, Y.; Nie, J.; Stewart, R.; Csordas, A.T.; Oh, S.S.; Thomson, J.A.; Soh, H.T. Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15373–15378. [[CrossRef](#)] [[PubMed](#)]
90. Schütze, T.; Wilhelm, B.; Greiner, N.; Braun, H.; Peter, F.; Mörl, M.; Erdmann, V.A.; Lehrach, H.; Konthur, Z.; Menger, M. Probing the SELEX process with next-generation sequencing. *PLoS ONE* **2011**, *6*, e29604.
91. Spiga, F.M.; Maietta, P.; Guiducci, C. More DNA-Aptamers for Small Drugs: A Capture-SELEX Coupled with Surface Plasmon Resonance and High-Throughput Sequencing. *ACS Comb. Sci.* **2015**, *17*, 326–333. [[CrossRef](#)] [[PubMed](#)]
92. Gotrik, M.R.; Feagin, T.A.; Csordas, A.T.; Nakamoto, M.A.; Soh, H.T. Advancements in Aptamer Discovery Technologies. *Acc. Chem. Res.* **2016**, *49*, 1903–1910. [[CrossRef](#)] [[PubMed](#)]
93. Ditzler, M.A.; Lange, M.J.; Bose, D.; Bottoms, C.A.; Virkler, K.F.; Sawyer, A.W.; Whatley, A.S.; Spollen, W.; Givan, S.A.; Burke, D.H. High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Res.* **2013**, *41*, 1873–1884. [[CrossRef](#)] [[PubMed](#)]

94. Scoville, D.J.; Uhm, T.K.; Shallcross, J.A.; Whelan, R.J. Selection of DNA Aptamers for Ovarian Cancer Biomarker CA125 Using One-Pot SELEX and High-Throughput Sequencing. *J. Nucleic Acids* **2017**, *2017*, 9879135. [[CrossRef](#)] [[PubMed](#)]
95. Hoon, S.; Zhou, B.; Janda, K.D.; Brenner, S.; Scolnick, J. Aptamer selection by high-throughput sequencing and informatic analysis. *Biotechniques* **2011**, *51*, 413–416. [[CrossRef](#)] [[PubMed](#)]
96. Kurtz, S.; Narechania, A.; Stein, J.C.; Ware, D. A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. *BMC Genom.* **2008**, *9*, 517. [[CrossRef](#)] [[PubMed](#)]
97. Weese, D.; Emde, A.K.; Rausch, T.; Doring, A.; Reinert, K. RazerS—fast read mapping with sensitivity control. *Genome Res.* **2009**, *19*, 1646–1654. [[CrossRef](#)] [[PubMed](#)]
98. Jing, M.; Bowser, M.T. Tracking the emergence of high affinity aptamers for rhVEGF165 during capillary electrophoresis-systematic evolution of ligands by exponential enrichment using high throughput sequencing. *Anal. Chem.* **2013**, *85*, 10761–10770. [[CrossRef](#)] [[PubMed](#)]
99. Thiel, W.H.; Giangrande, P.H. Analyzing HT-SELEX data with the Galaxy Project tools—A web based bioinformatics platform for biomedical research. *Methods* **2016**, *97*, 3–10. [[CrossRef](#)] [[PubMed](#)]
100. Thiel, W.H. Galaxy Workflows for Web-based Bioinformatics Analysis of Aptamer High-throughput Sequencing Data. *Mol. Ther. Nucleic Acids* **2016**, *5*, e345. [[CrossRef](#)] [[PubMed](#)]
101. Alam, K.K.; Chang, J.L.; Burke, D.H. FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Mol. Ther. Nucleic Acids.* **2015**, *4*, e230. [[CrossRef](#)] [[PubMed](#)]
102. Ducongé, F. Improvement of aptamers using PATTERNITY-seq (high-throughput analysis of sequence patterns and paternity relationship between them). In Proceedings of the the Aptamer Bordeaux Conference, Bordeaux, France, 24–25 June 2016.
103. Nguyen Quang, N.; Perret, G.; Duconge, F. Applications of High-Throughput Sequencing for In Vitro Selection and Characterization of Aptamers. *Pharmaceuticals* **2016**, *9*, 76. [[CrossRef](#)] [[PubMed](#)]
104. Hoinka, J.; Berezhnoy, A.; Sauna, Z.E.; Gilboa, E.; Przytycka, T.M. AptaCluster—A Method to Cluster HT-SELEX Aptamer Pools and Lessons from its Application. *Res. Comput. Mol. Biol.* **2014**, *8394*, 115–128. [[PubMed](#)]
105. Hoinka, J.; Dao, P.; Przytycka, T.M. AptaGUI-A Graphical User Interface for the Efficient Analysis of HT-SELEX Data. *Mol. Ther. Nucleic Acids.* **2015**, *4*, e257. [[CrossRef](#)] [[PubMed](#)]
106. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)] [[PubMed](#)]
107. Dao, P.; Hoinka, J.; Takahashi, M.; Zhou, J.; Ho, M.; Wang, Y.; Costa, F.; Rossi, J.J.; Backofen, R.; Burnett, J. AptaTRACE Elucidates RNA Sequence-Structure Motifs from Selection Trends in HT-SELEX Experiments. *Cell Syst.* **2016**, *3*, 62–70. [[CrossRef](#)] [[PubMed](#)]
108. Ding, Y.; Chan, C.Y.; Lawrence, C.E. S fold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* **2004**, *32*, W135–W141. [[CrossRef](#)] [[PubMed](#)]
109. Chan, C.Y.; Lawrence, C.E.; Ding, Y. Structure clustering features on the S fold Web server. *Bioinformatics* **2005**, *21*, 3926–3928. [[CrossRef](#)] [[PubMed](#)]
110. Hofacker, I.L. RNA secondary structure analysis using the Vienna RNA package. In *Current Protoc Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009.
111. Wang, J.; Gong, Q.; Maheshwari, N.; Eisenstein, M.; Arcila, M.L.; Kosik, K.S.; Soh, H.T. Particle Display: A Quantitative Screening Method for Generating High-Affinity Aptamers. *Angew. Chem. Int. Ed.* **2014**, *53*, 4796–4801. [[CrossRef](#)] [[PubMed](#)]
112. Bowser, M.T. SELEX: Just another separation? *Analyst* **2005**, *130*, 128–130. [[CrossRef](#)] [[PubMed](#)]
113. Katilius, E.; Flores, C.; Woodbury, N.W. Exploring the sequence space of a DNA aptamer using microarrays. *Nucleic Acids Res.* **2007**, *35*, 7626–7635. [[CrossRef](#)] [[PubMed](#)]
114. Platt, M.; Rowe, W.; Knowles, J.; Day, P.J.; Kell, D.B. Analysis of aptamer sequence activity relationships. *Integr. Biol.* **2009**, *1*, 116–122. [[CrossRef](#)] [[PubMed](#)]
115. Knight, C.G.; Platt, M.; Rowe, W.; Wedge, D.C.; Khan, F.; Day, P.J.; McShea, A.; Knowles, J.; Kell, D.B. Array-based evolution of DNA aptamers allows modelling of an explicit sequence-fitness landscape. *Nucleic Acids Res.* **2009**, *37*, e6. [[CrossRef](#)] [[PubMed](#)]
116. Rowe, W.; Platt, M.; Wedge, D.C.; Day, P.J.; Kell, D.B.; Knowles, J. Convergent evolution to an aptamer observed in small populations on DNA microarrays. *Phys. Biol.* **2010**, *7*, 036007. [[CrossRef](#)] [[PubMed](#)]

117. Nonaka, Y.; Yoshida, W.; Abe, K.; Ferri, S.; Schulze, H.; Bachmann, T.T.; Ikebukuro, K. Affinity improvement of a VEGF aptamer by in silico maturation for a sensitive VEGF-detection system. *Anal. Chem.* **2012**, *85*, 1132–1137. [[CrossRef](#)] [[PubMed](#)]
118. Kinghorn, A.B.; Dirkzwager, R.M.; Liang, S.; Cheung, Y.-W.; Fraser, L.A.; Shiu, S.C.-C.; Tang, M.S.; Tanner, J.A. Aptamer Affinity Maturation by Resampling and Microarray Selection. *Anal. Chem.* **2016**, *88*, 6981–6985. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).