Nucleic Acid Selection and the Challenge of Combinatorial Chemistry

Scott E. Osborne and Andrew D. Ellington*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received September 18, 1996 (Revised Manuscript Received December 10, 1996)

Contents

I. Introduction

Drugs are small organic molecules. This axiom has dominated the thinking of both chemical academics and the pharmaceutical industry for decades, and for good reason: it works. Small organic compounds can be made relatively cheaply in large yields and have proven to be extremely effective at manipulating the biochemistry and physiology of humans. The effectiveness of small organic compounds is in turn due in large part to their ability to fit tightly into the active sites or regulatory domains of enzymes, the binding pockets of receptors, and the interfaces of macromolecular complexes.

Drugs can be designed or found or both. Designing new drugs generally relies on an intimate understanding of the biochemistry of a disease state, including in many cases a knowledge of the structure of a biopolymer target. In contrast, finding drugs involves screening compounds with assays that are indicative of a disease state. Until recently, the organic compounds that were screened were by and large natural products and their derivatives. However, synthetic chemists have more recently found that they can rival Nature's bounty by developing methodologies that allow the modular synthesis of compounds.[1,2] Multiple reagents of a given type (e.g., primary amines) are added either together or in parallel at different steps of a synthetic scheme, resulting in a "combinatorial" accumulation of compounds. As before, these combinatorial chemical libraries are screened using automated, high-throughput assays, and the best leads are subjected to further modification and testing.

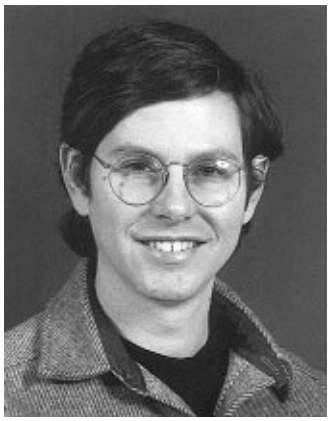
While the current paradigm for drug development is the identification of small organic effectors of metabolism, it is not the only possible route. Biopolymers are also extremely adept at altering the function of organisms, organs, cells, or, ultimately, other biopolymers. The utility of biopolymer drugs is attested by the fact that proteins such as insulin and erythropoetin have generated markets that are in the billions of dollars. Biopolymer drugs can be identified by design or screening, but there is also an additional route to their discovery: selection. Random sequence mixtures of peptides, proteins, or nucleic acids can be generated and individual molecules winnowed from the population by allowing them to bind to or be utilized by an enzyme, receptor, or cellular target. The selected molecules are then amplified *in vitro* or *in vivo*. Multiple cycles of selection and amplification will foster competition between active compounds, and should eventually result in the purification of those few molecular species that have the highest affinity or efficacy for a given target (Figure 1). In contrast, most organic compounds are not readily replicable and therefore cannot be selected and evolve in the same way that biopolymers can. It should be noted, however, that organic chemists have begun to blur the distinction between replicable biopolymers and nonreplicable organic molecules. For example, chemical libraries can be prepared in which information regarding the binding activity or efficacy of an individual compound or set of compounds is linked to sequence (e.g., position scanning libraries or SURF),[3,4] locale (VLSIPS),[5] or an encoded replicable or non-replicable tag.[6−8] Each method allows sublibraries with a desired activity to be resynthesized and rescreened. Chemical libraries in effect become "replicable" and can be selected through the investment of organic synthetic effort rather than biosynthetic transformations.

In order for biopolymer selection to be recognized as an important contender in drug discovery efforts, especially in light of the great strides that have been made in combinatorial chemistry, one of two conditions must be recognized: either selected biopolymers must become drugs that have unique properties or are inherently better than small organics, or selected biopolymers must be useful for the generation of conventional pharmaceuticals. Using the *in vitro* selection of nucleic acids as an example, we seek to establish why both conditions are likely.

While there have been numerous reviews of both the methods for and the products of *in vitro* selection,[9−15] we will approach the subject from a more chemical perspective. A brief review of the

S0009-2665(96)00009-X CCC: $28.00

Scott E. Osborne was born in Lebanon, IN in 1968. While majoring in chemistry and lettering three years in collegiate golf at DePauw University in Greencastle, Indiana, he conducted independent research under Professor Bryan Hanson. After receiving his B.A. in 1990, he attended the University of Michigan to pursue a doctoral degree in chemistry. Under the supervision of Professor Gary D. Glick, he developed a method to conformationally restrain oligonucleotides using disulfide cross-links. His efforts in this area were recently recognized with the biennial presentation of the Kasimir Fajans award for outstanding dissertation. Since receiving his Ph.D. in 1996, Dr. Osborne has been performing postdoctoral research under the supervision of Professor Andrew D. Ellington. His research interest focuses on using the chemistry developed at Michigan to stabilize small oligonucleotides that bind specific ligands (minimal aptamers) as well as probe the conformations of unique nucleic acid:protein complexes. He enjoys spending his spare time with his wife, Amy, and son, Trevor, as well as working on his golf and bridge games.



Andrew D. Ellington was born in Independence, MO in 1959. He graduated from Michigan State University with a B.S. in Biochemistry in 1981, and from Harvard with a Ph.D. in Biochemistry and Molecular Biology in 1987. His graduate work with Professor Steven Benner explored the role of point mutations in protein evolution and was carried out both at Harvard and at the Eidgenosische Technische Hochschule in Zurich, Switzerland. During this time, he became interested in the early stages of molecular evolution and postulated interconnections between the RNA world hypothesis and intermediary metabolism. In order to test this hypothesis, he carried out postdoctoral research with Dr. Jack Szostak at Massachusetts General Hospital and developed a method for the selection of functional nucleic acids. As a faculty member in the Chemistry Department at Indiana University, he has expanded on these results and pursued interests ranging from origins of life chemistry to antiviral gene therapies. Minimal spare time is spent petting his dogs.

method will yield principles of experimental design that are generally relevant to combinatorial chemistry. The specificities and affinities of selected nucleic acids for their targets will be examined in order to determine if nucleic acids enjoy some special advantage over organic compounds. Conversely, we will recount how the two chief disadvantages of potential nucleic acid pharmaceuticals, size and instability, may be remedied by appropriate chemical modifications. Next, the implications of selection experiments for conventional drug design will be

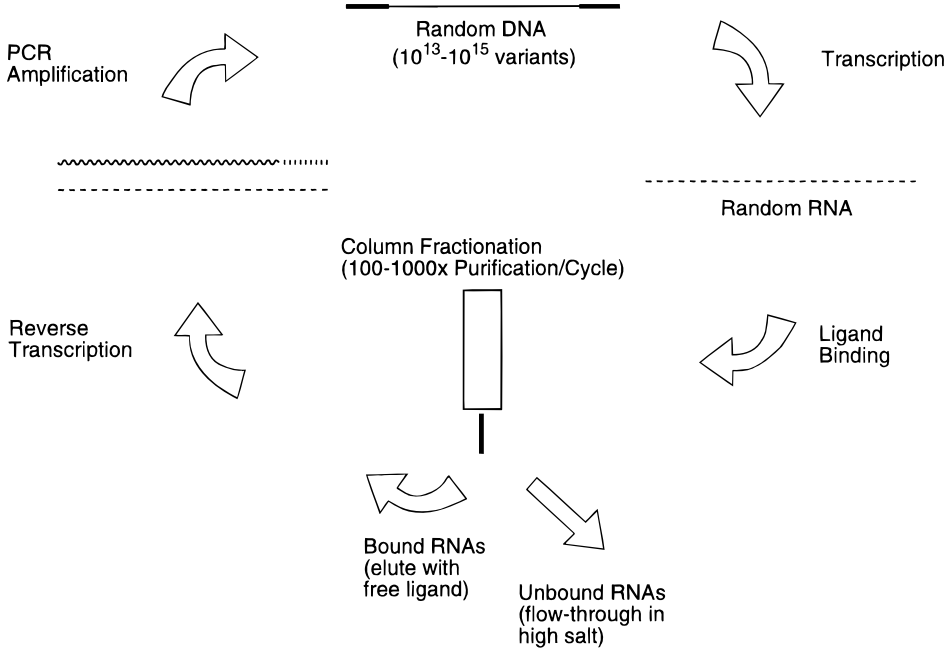


Figure 1. General schema for *in vitro* selection. A random sequence, single-stranded DNA pool is generated by chemical synthesis. Constant sequences flanking the random core are used for enzymatic manipulation. For example, the 5′-end might contain a T7 promoter and second strand priming site, while the 3′-end might contain a cDNA priming site (black boxes). Transcription with a RNA polymerase generates random sequence RNA, which can then be subjected to affinity purification (the selection process). Those RNAs that do not bind the immobilized ligand are washed away and discarded, whereas those which favorably interact with the ligand are selectively retained. Following affinity elution, the binding species are reverse transcribed to cDNA, amplified by PCR, and transcribed into RNA. The pool of random oligonucleotides is then once again subjected to the selection process. Multiple cycles of *in vitro* selection will result in selective purification of sequences that have a high affinity for the ligand of choice.

detailed. We argue that selection can be used both to define nucleic acid targets and to hone the specificities of compounds that bind to and inhibit the functions of nucleic acids.

II. A Brief History of Nucleic Acid Selection

It is frequently assumed and sometimes claimed on the basis of several arresting publications that the *in vitro* selection of nucleic acids was invented circa 1990. In fact, there is a large amount of prior art that precedes and presumes the development of modern methods for *in vitro* selection. The realization that nucleic acids could embody structure and function as well as carry information immediately suggested to experimentalists that it should be possible to carry out selection experiments at the molecular level. In the absence of a collection of purified, commercially available enzymes and amplification techniques such as the polymerase chain reaction (PCR), Sol Spiegelman and his co-workers used small RNAs derived from the Qβ bacteriophage genome and an encoded polymerase, the Qβ replicase, to develop a system for the *in vitro* replication and evolution of molecules.[16] This system accurately mimicked many of the properties of natural selection at the macroscopic level: the "fit" (or the fecund) survived, competition resulted in changes in the frequencies of different genotypes, mutants with novel properties emerged, parasites preyed on replicators.[17−19] Although the first *in vitro* selection experiments were performed over 20 years ago they nonetheless provided an important intellectual framework for what was to become a rapidly expanding field. The development of automated methods for the chemical synthesis of DNA allowed extremely diverse, random sequence libraries to be constructed, while the invention of *in vitro* amplification techniques, such as the polymerase chain reaction, enabled researchers to efficiently recover selected sequences and carry out multiple cycles of selection and amplification. Together, these advances prompted researchers to expand *in vitro* selection experiments beyond the molecular biology of the Qβ phage. A plethora of different selection experiments were carried out independently and nearly simultaneously, a fact which belies the notion that a single epiphany was responsible for the development of the field. Kevin Struhl, Larry Loeb, and their co-workers were among the first to take advantage of random synthetic libraries to probe oligonucleotide function, and selected functional promoter sequences either *in vitro* or *in vivo*.[20−23] As the use of random sequence libraries became increasingly popular, a number of researchers also defined binding sites for regulatory proteins.[24−28] The extrapolation of *in vitro* selection techniques from double-stranded to single-stranded nucleic acid shapes was obvious and led to the accelerated discovery of selected ligands (sometimes called aptamers) for a wide variety of targets.[29,30] Today, *in vitro* selection (sometimes called SELEX, the Systematic Evolution of Ligands by EXponential enrichment[29]) has moved beyond its academic roots to encompass the pursuit of lead compounds for therapeutic applications and the design of new diagnostic assays.

III. How It Works

A review of the methods used in *in vitro* selection experiments will provide essential background for a comparison with combinatorial chemistry. The *in vitro* selection process begins with a pool of sequence and structural diversity. Early experiments relied on the fact that Qβ replicase is extremely error-prone and generates numerous mutations each time it reproduces an RNA template. However, the number of different sequences that can be generated enzymatically is dwarfed by the number of different sequences that can be manufactured chemically. By programming a DNA synthesizer to add equal amounts of G, A, T, and C phosphoramidites at each step of a chemical synthesis, populations of 10^{13} to 10^{16} different DNA oligonucleotides can be routinely created. The use of solid phase oligonucleotide synthesis also allows researchers to more precisely control other characteristics of a nucleic acid pool. The length of the random sequence tract can be programmed to be anywhere from 1 to up to 140 or more residues. The degree of randomization can also be varied by varying the composition of the phosphoramidite mixtures. While many selections start with completely random nucleic acid pools, natural nucleic acid ligands can be partially randomized to contain, say, 70% wild-type and 10% of each non-wild-type residue (i.e., if a position in a ligand were a guanosine the pool would contain 70% G, 10% A, 10% T, and 10% C at that position). Partial randomization generates a pool with multiple mutations centered on the wild-type sequence.[31] Finally, chemical synthesis allows details regarding how a nucleic acid pool will be manipulated to be specified in advance. Constant sequence regions typically flank the random sequence tracts and allow the nascent DNA oligomer to be amplified by PCR. Promoter sequences included within the constant regions facilitate the enzymatic transformation of a DNA pool into an RNA or even modified RNA pool.

Each member of the nucleic acid pool has a different sequence and, thus, a different set of chemical groups that will fold into differing structures that have differing properties or are capable of different functions. Individual members of a pool are separated from one another on the basis of their ability to perform an arbitrary task. For example, RNA and DNA molecules that can carry out catalytic reactions such as ligation and phosphorylation have been culled from random sequence mixtures.[32,33] The focus of this review is nucleic acid binding species (aptamers), and a wide variety of methods have been adapted to their selection. For example, nucleic acids that interact with protein targets have been trapped on nitrocellulose filters or separated from uncomplexed species on native polyacrylamide gels. Aptamers that bind small molecules such as cofactors or vitamins can be isolated by affinity chromatography. Unbound nucleic acids are washed away from bound, and the bound species are then eluted by adding an excess of soluble ligand, changing solvent conditions, or cleaving from the solid support (via a reversible disulfide cross-link, for example). As was the case for pool synthesis, the progress of an *in vitro* selection can be more precisely controlled than the progress

of a natural selection. The stringency of a selection can be varied by changing the concentration of target, the buffer conditions for binding, and the volume of wash that precedes elution.[13,14,34]

Up to this point, there is little difference between screening a chemical library and a nucleic acid library. In both cases, only a few binding species may remain in the population following the initial screen. However, nucleic acids can be directly amplified using conventional molecular biology methods. In early experiments, Qβ replicase provided up to 10^9-fold amplification of successful sequences. Unfortunately, Qβ replicase discriminates between sequences on the basis of their suitability as templates or their replicability, and thus might inadvertently amplify a survivor with medium affinity at the expense of a poorer replicator with higher affinity. Newer methods allow greater control over the process of amplification and yield fewer artifacts. Selected DNA molecules can be readily amplified using PCR. RNA or modified RNA molecules must first be converted to DNA copies using reverse transcriptase, but can then also be amplified with the PCR. RNA can be regenerated from the amplified DNA via *in vitro* transcription with a bacteriophage polymerase, such as those from phage SP6 or phage T7. Other amplification methods combine these steps; for example, the transcription-based amplification system (TAS, also known as 3SR)[35] utilizes reverse transcriptase and RNA polymerase to cycle between DNA and RNA intermediates, mimicking the life cycle of retroviruses.

Multiple cycles of selection and amplification result in competition between binding species and allow aptamers with the highest affinity to be extracted from a population. The number of cycles of selection required to preferentially amplify a binding species or set of species is a function of the ligand, degree of randomization, and stringency, but from three to 15 cycles is typical. The entire process can take from a few days to a few months, but on average 2−4 weeks suffices. After the population has been pared down to those sequences which bind the ligand with high affinity, the pool is cloned and individuals sequenced. The number of discrete sequences that remain in a population following selection will range from one to thousands. Comparisons between selected sequences and their predicted secondary structures frequently reveal similar motifs or substructures that are important for binding function.[36] However, it is equally likely that aptamers will fall into unrelated families; that is, there may be multiple sequence "answers" to a given binding "problem".

IV. What Selection Reveals about Combinatorial Chemistry

Superficially, results garnered from *in vitro* selection experiments may not seem applicable to the design and screening of chemical libraries. As we have emphasized, chemical libraries are not replicable, while nucleic acids, for all of their diversity, are not small organic molecules that can be readily converted to drugs. However, nucleic acid libraries are chemical combinatorial libraries in the broadest sense of the term, and in both cases the problem of sorting through large numbers of different sequences (or compound formulae) and structures remains the same. Therefore, it may be possible to apply some of the lessons learned from nucleic acid selections to chemical library screens. For example, some of the considerations encountered during the design and execution of *in vitro* selection experiments, such as the problems of library coverage and skewing, may have counterparts in screens for pharmacophores. Such comparisons may or may not eventually be shown to be accurate, but their true value lies in how they can aid researchers to develop principles for experimental design in the nascent field of combinatorial chemistry.

Table 1. Number of Different Oligonucleotides Based on the Number of Random Residues

no. of random residues	no. of sequentially different compounds
1	4
4	256
10	$\approx 10^6$
20	$\approx 10^{12}$
30	$\approx 10^{18}$
40	$\approx 10^{24}$

Neither nucleic acid libraries nor chemical libraries completely span the universe of all possible structures. Because there are only four possible (natural) monomers, it is easier to quantitate the diversity and completeness of nucleic acid libraries (Table 1). For example, a nucleic acid pool that has 40 randomized positions potentially contains 4^{40} or about 10^{24} different sequences, each one of which will have a unique associated structure. In practice, though, synthetic limitations confine most libraries to at most 10^{16} different sequences. Therefore, in this example only one in every 10^8 possible base combinations is represented. Of course, mutations that arise during the amplification process increase the total number of species that are examined during the course of the selection, and for this reason amplification procedures that are inherently mutagenic are sometimes used. Nonetheless, for most selection experiments the important statistic is not the total number of species nor the coverage of "sequence space", but rather the number of successful species that are originally present in the actual population. This ratio varies from approximately $1:10^6$ to $1:10^{12}$, depending on the target and the stringency of selection. In general, library size is not an obstacle to the identification of aptamers. Similarly, although combinatorial chemical libraries are frequently only 10^3 to 10^6 strong, they nonetheless contain an impressive array of compounds, many of which are found to be active in a variety of assays. For example, Ellman has identified numerous compounds with novel binding specificities from a chemical library based solely on relatively limited modifications of a benzodiazepine backbone.[37]

It is less clear whether or not these incomplete nucleic acid libraries or chemical libraries of necessity contain the best possible compounds. Of course, in the absence of carrying out experiments with libraries that completely span sequence or structure spaces, it may never be possible to know. Once again, although, some *in vitro* selection experiments provide

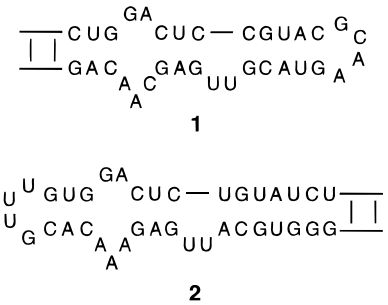


Figure 2. Consensus sequence of anti-Rev aptamers from nucleic acid libraries containing 18 (**1**)[38] and 32 (**2**)[39,40] random residues. In all cases throughout this review, the top stand of an oligonucleotide sequences reads 5′ to 3′.

clues as to whether incomplete libraries are "complete enough". RNA molecules that can bind to the Rev protein of HIV-1 have been selected from random sequence libraries of either 18 or 32 residues in length.[38−40] The library that contained 18 random sequence positions was likely complete, while the library that spanned 32 residues was not. Nonetheless, high-affinity aptamers with similar motifs were recovered from both libraries (**1** and **2**, Figure 2), indicating that at least in this instance the lack of coverage did not affect the outcome of the search. Similarly, aptamers that could bind to vascular endothelial growth factor (VegF) have been isolated from modified RNA libraries that spanned either 30 or 50 random sequence positions; the same high-affinity binding motif was recovered from both libraries.[41]

It may be possible to improve the probability of finding the best possible binding species by presenting random sequences within the context of a structured nucleic acid, just as antibodies present their variable loops within the immunoglobulin framework. For example, while Giver and co-workers isolated anti-Rev aptamers from a library in which up to 18 residues within an internal loop were randomized,[38] these authors also carried out a second selection starting from a library in which only 10 residues within an internal loop were randomized. The aptamers isolated from the two libraries had different sequences and binding characteristics: the aptamers isolated from the library with 18 randomized residues bound roughly 3-fold better than those isolated from the library with the shorter random sequence tract. These results can be interpreted to indicate either that more diverse libraries inherently contain better binding sequences or that the flanking constant sequences in the shorter library may have influenced the course of the selection. Similarly, Hamm synthesized nucleic acid libraries in which random sequence tracts were presented either in the context of a secondary structural elements, a G-quartet, or as an unstructured random sequence tract.[42] Aptamers that could bind to an antiferritin antibody were selected in parallel from both libraries. Unfortunately, the results of these experiments could not be interpreted to determine whether or not structure-specific libraries assist in identifying high-affinity binding species, since the primer binding sites in the constant regions paired with one another to form a strong secondary structure. This design resulted in the selection of species that could most efficiently bind primers and be amplified, rather than species that could bind to the antibody, and the selected populations were correspondingly skewed.

Overall, one possible interpretation of these experiments is that sequence motifs that are optimal for binding are frequent enough that they can be found even in incomplete libraries. Put another way, it may be either that long (and hence rare) sequence motifs are not uniquely suited to binding or that numerous variants of long sequence motifs (at least some of which will be found in an complete library) are more or less equally good at high-affinity binding. Translated from nucleic acid to chemical libraries, the implication is that complete coverage of functional space can be achieved even with limited libraries. Theoretical analyses by Peter Schuster and co-workers are consistent with these empirical observations.[43] These authors have demonstrated using a simple folding algorithm that even relatively limited sets of nucleic acid sequences can completely cover secondary structural space, making it unlikely that any but the rarest shapes will be underrepresented in a random sequence pool. However, there is an important caveat to this analysis: nucleic acids are linear polymers that fill shape and functional spaces by folding, and the folded structures have some inherent flexibility. Indeed, some selected nucleic acids, such as anti-nerve growth factor (NGF) aptamers,[44] only become structured on binding to their targets. Compounds within chemical libraries tend to be far more constrained, and hence may not represent the same dynamic ensemble of shapes. Instead, chemical libraries cover shape space by utilizing a much wider variety of fixed functional groups. Extending this comparison, it is possible that monomer diversity and monomer flexibility may compensate for one another in the design of chemical libraries: a smaller number of flexible monomers may be as useful for lead discovery as a larger number of more fixed monomers, despite the entropic cost that is paid on binding. This principle may explain why chemical libraries with some inherent flexibility, such as Chiron's peptoids, have proven successful.[45]

An alternative explanation for why selections that start with different pools give the same sequence answer is that the shortest answer typically wins. This phenomena has been called the "tyranny of short motifs",[46] and can be readily illustrated: in a random sequence pool, a particular 10 nucleotide motif will appear in roughly one in a million molecules, while a 20 nucleotide motif will appear in one in a trillion molecules, or a million-fold less frequently. Unless the 20 nucleotide motif has an affinity for a target molecule that is much higher than that of the 10 nucleotide motif, it will almost never be found in a selected population. While the inherent numerical advantage of shorter motifs can be partially offset by increasing the stringency of the selection, the highest affinity aptamers may never be known if small nucleic acid structures can form relatively tight complexes with their targets. The identification of short but suboptimal motifs has been observed in the selection of RNA molecules; several examples should suffice to link theory with experiment:

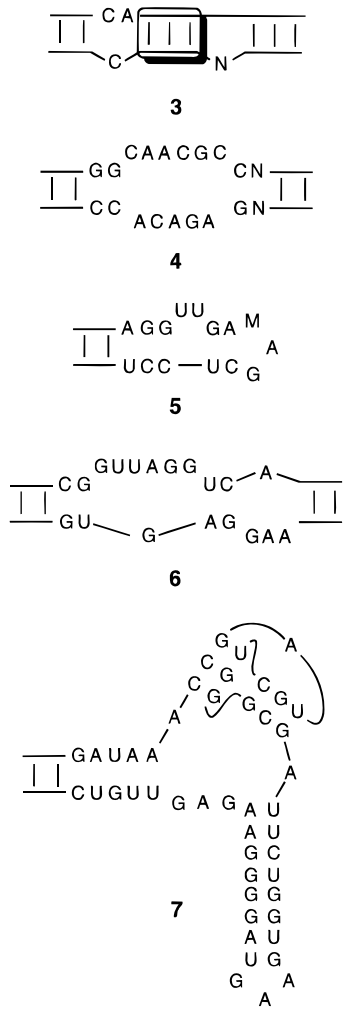


Figure 3. Consensus sequences of aptamers (**3**) that bind D-arginine nearly 2-fold better than L-arginine (box indicates that different canonical base pairs were conserved in 21 of 23 variants examined; N indicates any residue).[47] An aptamer was also recovered that could bind both arginine and guanosine (**4**). The aptamer found by Tao and Frankel (**5**) resembled a *TAR*-like structure (M indicates A, G, or U).[49] Famulok and co-workers derived two different aptamers (**6**)[50] and (**7**)[51] from long, random sequence pools.

(1) Selection of anti-arginine aptamers. Yarus and his co-workers have selected RNAs that can bind to arginine (**3**)[47] or to arginine and guanosine (**4**)[48] from random sequence pools (Figure 3). In both cases the random sequence regions were relatively short (25 residues), and the selected motifs utilized the constant flanking regions to form functional secondary structures; in other words, at least part of the information in the selected motif was "fixed" prior to selection. The dissociation constants for the arginine:RNA complexes ranged from 200 μM to 4 mM. Tao and Frankel also selected RNAs that could bind to arginine from a pool that spanned 30 random sequence positions.[49] While the recovered motifs (**5**) did not rely on constant sequences for binding function, the dissociation constants of the arginine:RNA complexes were again in the millimolar range. In contrast, Famulok selected anti-arginine aptamers (**6**) from a partially randomized pool that spanned 74 positions;[50] the dissociation constants for the arginine:RNA complexes were as low as 10 μM. The sequences of the selected aptamers bore no resemblance to those selected previously. This same group also selected anti-arginine aptamers from a completely randomized pool that spanned 113 positions;[51] the dissociation constants for the best arginine:RNA complexes were 330 nM. The sequences of the selected aptamers (**7**) bore no resemblance to those selected previously.

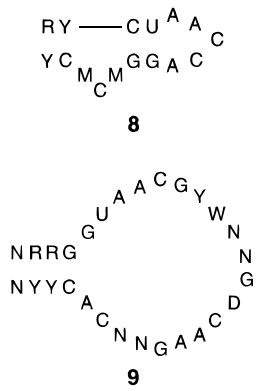


Figure 4. The two predominant aptamer families isolated from the selection against bFGF in the absence (**8**) and presence (**9**) of heparin (R = A or G; Y = C or U; M = A or C; W = A or U; D = A, G, or U; and N = any base).[52]

(2) A correlation between selection stringency and motif complexity was observed in selections that targeted basic fibroblast growth factor (bFGF).[52] Two RNA pools that spanned 30 random sequence positions were used as starting points for the isolation of anti-bFGF aptamers; the first pool was selected against bFGF alone, while the second pool was selected against bFGF in the presence of heparin. Two predominant aptamer families were isolated (Figure 4). Family 1 was predicted to form a stem-loop that contained a side bulge (**8**), while family 2 was predicted to form a stem capped by a very large loop (**9**). The first selection with the first pool returned both family 1 motif and family 2 motifs, while the second selection with the second pool returned primarily the family 2 motif. Both families interacted with the heparin binding site of bFGF, but family 2 aptamers bound roughly 10-fold better to bFGF than family 1 aptamers. While it may be that the differences in constant region sequences between the two pools favored the isolation of family 2 motifs in the second selection, the fact that this motif was found in several registers in selected aptamers argues against this. Rather, it may be that the inclusion of heparin favored the selection of family 2. This hypothesis was borne out by examining the complexity of the selected ligands. The "information content" of these families was estimated on the basis of the number of conserved and semiconserved residues. The family 1 motif was expected to occur once every 3.6×10^7 molecules, while the family 2 motif was expected to occur only once every 4.8×10^{11} molecules. Thus, the less stringent selection returned molecules of varying information content, while the more stringent selection returned the less probable motif. This is precisely the result that would have been predicted based on our understanding of the tyranny of short motifs.

(3) Ribozymes as well as aptamers appear to be subject to the "tyranny". Bartel and Szostak selected RNA ligases from a random sequence pool that

spanned 220 positions.[32] Three families of ligases were identified; two classes were structurally simple and contained multiple representatives, while the final class was extremely complex and contained only a single representative.[53,54] Again, this result is consonant with the tyranny of short motifs. Interestingly, the single, complex ribozyme should not have been found at all: its "information content" was so high that it should have been present only once in every 2.5×10^{18} sequences, while the original pool spanned only 1.4×10^{15} sequences. These results imply either that Bartel was extraordinarily lucky, or that multiple, complex motifs were present in the population. If the latter hypothesis is true, it may resolve a potential contradiction between the "tyranny" and the observation we made above that even incomplete libraries seem to be "complete enough" to successfully carry out selection experiments. While many complex, highly functional motifs may not be present in an incomplete library, if there are a number of these motifs at least one is likely to be found.

Taken together, these examples verify the reality of the "tyranny of short motifs," and further support the contention that longer, more complex pools can return aptamers that are as good as or better than those found in shorter, less complex pools, even when the latter are structurally constrained. In order to translate these results from RNA to chemical libraries, it is necessary to assume that the structural and functional diversity generated by sequence is equivalent to the structural and functional diversity generated by differing chemical groups. If this equivalence is true, the more structurally diverse the chemical library, the better the likelihood of selecting a high-affinity drug candidate. While this hypothesis currently forms the basis for the design of numerous chemical libraries, empirical support for it has been difficult to obtain.

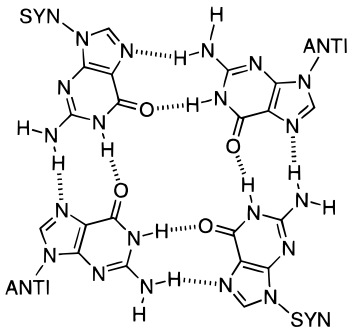
V. The Advantages of Selected Nucleic Acids

While organic compounds can interact tightly and specifically with target molecules, biopolymer ligands will frequently exhibit even better binding characteristics. This generalization is rooted in the fact that large "hosts" and large "guests" can form large interfaces. The best drugs will bind in clefts or crevices on a target molecule, forming an intricate network of interactions. The best biopolymers can also bind in clefts or crevices, but will in addition generate clefts or crevices of their own that can enfold surface features of the target. The more extensive interaction surface provides additional opportunities for a biopolymer "host" to discriminate between closely related target "guests". This generalization is also bolstered by the exquisite specificities of known antibody:antigen interactions. Consider that antibodies cannot only bind to surface features of protein antigens, but can actually induce conformational changes in the epitopes that are recognized. As the structure of the interface between antibody and antigen deforms, each molecule has several chances to reject any noncognate partner.

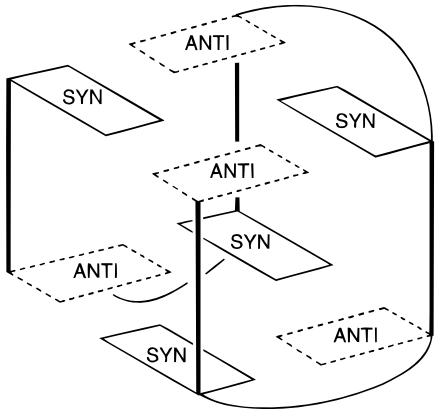
There are multiple examples of how selected nucleic acids can discriminate between even closely related targets. For example, aptamers selected to bind one reverse transcriptase (RT) do not interact with related proteins. RNA ligands to RT from the human immunodeficiency virus (HIV-1) did not inhibit similar enzymes from avian myeloblastosis virus (AMV) or Moloney murine leukemia virus (MMLV). RNA ligands to feline immunodeficiency virus RT did not inhibit similar enzymes from AMV, MMLV, and HIV-1.[55] Anti-bFGF aptamers bound to their cognate cytokine from 100- to 1000000-fold better than related fibroblast growth factors.[56] RNA aptamers directed against the βII isozyme of protein kinase C did not inhibit the α isozyme,[57] which is 80% homologous, and strongly discriminated against the βI isozyme, which differs by only 37 of 673 residues localized primarily in the carboxy terminus. DNA aptamers to human thrombin showed preferential binding to thrombin over other serine proteases.[58] The anti-thrombin aptamer could even discriminate against protein variants that differed by only a single amino acid. A mutation (R70E) in thrombin exosite I eliminated interactions with the thrombin aptamer.[59] Interestingly, DNAs selected to bind to the R70E variant could interact with wild-type thrombin as well,[60] indicating that there may be aptamer "generalists" as well as "specialists".

As expected, the basis for aptamer selectivities appears to be their ability to meld with surfaces features of a protein, fitting to crevices and/or enfolding protrusions. Aptamers selected to bind to the human T-cell leukemia virus regulatory protein Tax interacted with a particular subset of functional features, blocking protein:protein interactions with some transcription factors but not others.[61] Aptamers selected to bind to Qβ replicase formed two classes that did not compete with one another for binding and recognized independent sites on the protein.[62] The interaction of aptamers with particular protein surface features has been impressively validated by structural studies of the anti-thrombin aptamer. The DNA 15-mer is found to form a G-quartet stack similar to that observed for telomeric DNA (Figure 5).[63−66] Mutational and chemical protection analysis had indicated that the aptamer interacted with the anion-binding exosite of thrombin,[59,67] and this was confirmed by a crystallographic study.[68] A loop on the thrombin aptamer pokes into the exosite, and phosphate oxygens contact basic amino acids while thymines are involved in hydrophobic interactions.

Aptamers can also wrap around amino acid sequence tracts, similar to the way antibodies bind to continuous epitopes on proteins. Nieuwlandt and coworkers selected aptamers that could bind to a peptide, substance P.[69] The aptamers could efficiently recognize oligopeptides as short as 7 amino acids in length, but did not recognize the sequence presented in a reverse orientation. Xu and Ellington selected aptamers that could bind to a 17-mer α-helical peptide derived from the Rev protein.[70] Binding studies with peptide variants indicated that some aptamers recognized individual amino acids. The aptamers could also recognize the same amino acid sequence when presented within the context of the Rev protein, just as antipeptide antibodies can frequently cross-recognize proteins.



10



11

Figure 5. Base-pairing scheme of G-quartet (**10**) in which the bases alternate between a *syn* and *anti* torsion angle in order to form a web of hydrogen bonds. Base-stacking arrangement of G-quartets found in the anti-thrombin aptamer (**11**) also alternate from *syn* to *anti*.[68]

The affinities of selected nucleic acids for their targets are generally quite good, with dissociation constants falling in the subnanomolar to decananomolar range. These values are comparable to those of conventional drugs and of some monoclonal antibodies or Fabs. The relative binding abilities of aptamers have previously been reviewed (Table 2).[9,12]

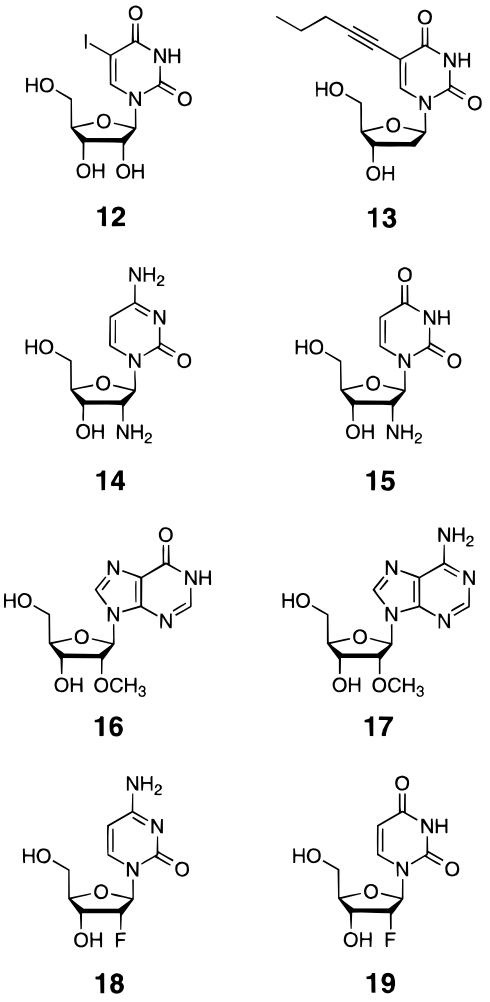
Recently, several methods have been developed to enhance the affinities of selected nucleic acids for their targets. Small molecules have been appended to aptamers in order to direct them to bind at particular sites on a target. In a variation of this method, Lin and co-workers tethered an aptamer selected to bind human neutrophil elastase to a tetrapeptide that was a weak inhibitor of the enzyme.[71] The two compounds bound at separate sites on the enzyme surface, and thus could act cooperatively when linked together. The affinity of the aptamer for elastase was improved by a modest 2-fold ($K_D \approx 7$ nM), but the inhibitory ability of the peptide increased by five orders of magnitude. In a separate set of studies, Smith and co-workers conjugated a RNA pool to a suicide inhibitor of elastase and selected RNA conjugates that could become covalently bound to the target.[72] Again, the small molecule increased the (noncovalent) binding affinity of the aptamer by around 20-fold. More impressively, conjugation reduced cross-reactivity with related targets by over 100-fold. Nucleic acids that can covalently cross-link to their targets can also be selected on the basis of chemistry inherent in the nucleic acid, rather than on the chemistry inherent in the target or its active site.[73] A nucleic acid pool was synthesized with a photolabile nucleotide, 5-iodouridine (**12**), in place of uridine, mixed with a protein target, HIV-1 Rev, and then exposed to UV light (Chart 1). Those nucleic acid species that bound Rev and positioned the photolabile nucleotide adjacent to a chemically compatible surface feature of the protein became cross-linked. The covalently bound species were isolated, amplified, and further selected both for their ability to bind in the absence of irradiation and for their ability to form extremely tight complexes that cannot be dissociated by denaturation. The ligands that were eventually discovered could bind with subnanomolar affinities and efficiently cross-linked to the target protein. Amazingly, some of the selected species also had the ability to cross-link even in the absence of irradiation. One of the many nucleophiles on the surface of the protein likely catalyzed the displacement of iodine and the formation of a covalent bond to the 5 position of uridine.

Table 2. Relative Binding Abilities of Aptamers to Various Proteins

ligand	aptamer motif	conjugate ligand K_D (nM)	ref
Nucleic Acid Binding Proteins			
AMV RT	hairpin	2	83
ASF/SF2, SC35	hairpin	N/A[a]	155
E. coli 30S Particles + S1	pseudoknot	5	156
E. coli 30S Particles - S1	Shine-Dalgarno	7	156
E. coli rho Factor	hairpin	1	*b*
E. coli S1 Protein	pseudoknot	4	156
FIV RT	hairpin	2	55
Hel-N1	U-rich	N/A	*c*
HIV-1 Int	complex	20	*d*
HIV-1 Rev	bulge	1	38−40
HIV-1 RT	pseudoknot	5	83
HIV-1 Tat	hairpin/bulge	5	39
HTLV-1 Rex	bulge	N/A	147
MMLV RT	hairpin	10	83
Qβ replicase	pseudoknot	5	62
R17 coat protein	hairpin	5	*e*
ribosomal protein L22	hairpin	N/A	*f*
T4 DNA polymerase	hairpin	5	29
U1A	hairpin	5	*g*
U2AF	single strand	1	154
Non-Nucleic Acid Binding proteins			
anti-gp10 antibodies	hairpin	N/A	*h*
anti-insulin receptor Ab	hairpin/bulge	2	*i*
bFGF	hairpin	0.2	52
HTLV-1 Tax	hairpin	100	61
NGF	pseudoknot	100	44
PKC bII	complex	7	57
substance P	hairpin/bulge	190	69
thrombin	hairpin	9	157
VegF	hairpin/bulge	0.2	126

[a] Not available or determined. [b] Schneider, D.; Gold, L.; Platt, T. *FASEB J.* **1993**, *7*, 201. [c] Levine, T. D.; Gao, F.; King, P. H.; Andrews, L. G.; Keene, J. D. *Mol. Cell. Biol.* **1993**, *13*, 3494. [d] Allen, P.; Worland, S.; Gold, L. *Virology* **1995**, *209*, 327. [e] Schneider, D.; Tuerk, C.; Gold, L. *J. Mol. Biol.* **1992**, *228*, 862. [f] Dobbelstein, M.; Shenk, T. *J. Virol.* **1995**, *69*, 8027. [g] Tsai, D. E.; Harper, D. S.; Keene, J. D. *Nucleic Acids Res.* **1991**, *19*, 4931. [h] Tsai, D. E.; Kenan, D. J.; Keene, J. D. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8864. [i] Doudna, J. A.; Cech, T. R.; Sullenger, B. A. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2355.

Chart 1. Structure of Modified Nucleotides Used To Increase Nuclease Resistance and Chemical Stability of Aptamers



VI. Making Nucleic Acids More Stable and Smaller

Any advantages selected nucleic acids may have in terms of affinity and specificity are obviated if they are too fragile for clinical use or too expensive to make. Unfortunately, RNA can be quickly degraded by a wide variety of nucleases found in sera and cells. In fact, the half-life of most natural or selected RNA molecules in sera is minutes or less. DNA molecules are slightly more stable, with half-lives up to hours. In addition, the development of antisense technology has collaterally yielded a variety of methods for the large scale synthesis and *in vivo* stabilization of DNAs. Thus, it is important to realize that *in vitro* selection experiments can be carried out with DNA libraries as well as with RNA libraries (Table 3). The first example, already described, of a DNA selection against a protein target was the identification of anti-thrombin aptamers.[58] While the initial selections returned a small, readily manipulable 15-mer motif, additional selections have found that this motif can be expanded to include an adjacent stem structure.[74] Anti-IgE aptamers were selected from a random sequence DNA pool that spanned 40 or 60 positions.[178] The aptamers bound their target with an affinity of 10 nM. Anti-reverse transcriptase aptamers have been selected from a random sequence DNA pool, bound the enzyme with a K_D of about 2 nM, and inhibited reverse transcriptase activity with a K_i as low as 0.3 nM.[75] As was the case with aptamers selected from RNA libraries, the selected DNA was specific for its cognate reverse transcriptase.

Table 3. Aptamer Selections Carried out Using DNA Pools

ligand	pool	aptamer motif	conjugate ligand K_D (nM)	ref
E. coli metJ protein	dsDNA	unknown	1	*a*
elastase	ssDNA	G-quartet	7	71
HIV-1 RT	ssDNA	stem/bulge	2	75
IgE	ssDNA	stem/bulge	10	178
Lrp	dsDNA	duplex	2	*b*
thrombin	ssDNA	G-quartet	25	58

[a] He, Y.; Stockley, P. G.; Gold, L. *J. Mol. Biol.* **1996**, *255*, 55. [b] Cui, Y.; Wang, Q.; Stormo, G. D.; Calvo, J. M. *J. Bacteriology* **1995**, *177*, 4872.

Table 4. Aptamer Selections Carried out Using Modified DNA or Modified RNA Pools

ligand	modified bases in pool[a]	aptamer motif	conjugate ligand K_D (nM)	ref
bFGF	**14**−**15**	unknown	0.3	78
elastase	**14**−**15**	G-quartet	15	77
HIV-1 Rev	**12**	hairpin/bulge	1	73
IgE	**14**−**15**	G-quartet	30	178
RSV	**18**−**19**	various	3 μg[b]	81
thrombin	**13**	hairpin	400	76
VegF	**14**−**17**	hairpin	0.3	41

[a] See Chart 1 for structures. [b] Expressed in terms of total viral protein concetration instead of K_D.

A. Modified Nucleotides

Further stabilization can be afforded by the introduction of chemical modifications into selected nucleic acids (Table 4). In general, modifications of the 2′ position of ribotides or of the phosphodiester backbone interfere with the enzymatic mechanisms of nucleases and lead to concomitant increases in stability. Modified nucleotides can be introduced either into the initial pool (presubstitution) or following selection of binding species (postsubstitution) or both.

Presubstitution strategies rely on the ability of DNA polymerases, or RNA polymerases and reverse transcriptases, to utilize modified nucleotides as substrates.[76,77] Several selections have now been carried out in which one or more of the canonical nucleotides in the original library was replaced with a modified nucleotide (Chart 1). In the first reported use of modified bases during selection anti-thrombin aptamers were isolated from a DNA pool that contained 5-(1-pentynyl)-2′-deoxyuridine (**13**).[76] This modified nucleotide had previously been shown to enhance the activities of antisense oligonucleotides and could be readily incorporated into growing strands by a thermostable polymerase. The anti-thrombin aptamers were dependent on the modified base for activity, but actually had less affinity for thrombin than aptamers selected from a "natural" DNA library. The modified nucleotides 2′-aminocytidine (**14**) and 2′-aminouridine (**15**) interfere with ribonuclease cleav-

age, are recognized by T7 RNA polymerase and AMV reverse transcriptase, and have been substituted for their natural counterparts in a number of selections. Lin and co-workers selected nucleic acid ligands that could bind to human neutrophil elastase and were completely substituted with these modified pyrimidines.[77] The aptamers were dependent on the amino sugars for high-affinity binding, and the half-lives of the substituted RNAs were increased to many hours in serum. Similarly, aptamers that bound bFGF or vascular endothelial growth factor (VegF) were selected from pools containing modified pyrimidines.[41,78] The anti-bFGF aptamers had K_D values as low as 3.5×10^{-10} M and their stability in serum was increased by at least 1000-fold relative to unsubstituted RNAs (other experiments suggest that the relative stabilities of such substituted RNAs is likely increased by 10^6-fold[79]). Sequence and deletion analyses of the anti-VegF aptamers allowed a minimal binding species (32 residues) to be designed. In order to further stabilize the construct to exonuclease degradation, several residues containing phosphorothioates were added to the 5′- and 3′-termini. The minimal aptamer containing modified nucleotides and phosphorothioate caps was found to form a complex with VegF that had a dissociation constant of 2.4 nM and its half-life in rat urine (an even more hostile environment than serum[41]) was 17 h.

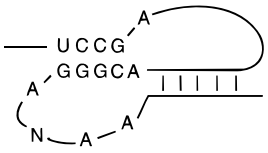
Because the identity of individual residues in selected sequences is critical for binding it would not necessarily be predicted that postsubstitution of modified nucleotides would lead to ligands that were both stable and efficacious. Surprisingly, modified nucleotides can be used to improve both properties. The anti-VegF aptamers that contained 2′-aminopyrimidines were further substituted with 2′-*O*-methylpurines (**16** and **17**).[41] Complete substitution of all purines in the minimal motif led to a drastic (100-fold) decrease in binding affinity. However, selected substitution of 10 of the 14 purines present actually led to a 10-fold increase in binding affinity and an 8-fold increase in stability; the final aptamer had a half-life in urine of 131 h. The improvement in affinity was largely due to a decrease in the dissociation rate of the complex. Green and co-workers were also able to use to postsubstitution to radically improve the stability of aptamers selected from RNA pools.[80] A novel interference analysis was used to determine which residues in an anti-reverse transcriptase aptamer could be substituted with 2′-*O*-methyl nucleotides. The aptamer was synthesized with a 1:1 mixture of normal nucleotides and 2′-*O*-methyl nucleotides, the population was sieved based on the basis of its ability to bind to HIV-1 RT, the bound species were hydrolyzed with base, and the mixture was analyzed by gel electrophoresis. When the incorporation of a 2′-*O*-methyl nucleotide at a given position interfered with binding, the selected population would contain predominantly normal ribotides at that position. Such "critical" positions were identified on the basis of their increased susceptibility to alkaline cleavage. Only two positions were found to require a 2′-hydroxyl moiety for high-affinity interactions with reverse transcriptase, the remainder of the ligand could be completely substituted with 2′-*O*-methyl nucleotides. Selection can also be used to facilitate postincorporation of modified nucleotides. Aptamers that could bind to Rous sarcoma virus (RSV) particles were selected from a random sequence RNA pool.[81] When aptamers from the selected pool were completely substituted with 2′-fluoropyrimidines (**18** and **19**), much of their binding activity was lost. Therefore, the selected pool was substituted with modified nucleotides and further selected for an additional three cycles for binding to RSV. The final pool containing modified pyrimidines could interfere with viral replication almost as well as the unmodified, parental population, indicating either that sequences that were dependent on unsubstituted pyrimidine nucleotides were eliminated from the population, or that aptamers containing positions that required unsubstituted pyrimidine nucleotides had mutated.

Given the chemical differences between RNA, modified RNA, and DNA it is perhaps not surprising that selections that start with different pools return different aptamers. Anti-RT aptamers selected from RNA pools have different sequences and structures than anti-RT aptamers selected from DNA pools; the same is true for anti-thrombin aptamers and for anti-ATP aptamers.[82] Similarly, anti-FGF and anti-VegF aptamers selected from RNA pools have different sequences and structures than the corresponding aptamers selected from modified RNA pools containing 2′-aminopyrimidines (**14** and **15**). Finally, anti-elastase aptamers selected from DNA pools have different sequences and structures than the corresponding aptamers selected from RNA pools containing 2′-aminopyrimidines or conjugated to a suicide inhibitor. These results may appear inconsistent with the fact that postsubstitution strategies for stabilization are successful. If most of the residues and the structural context of a selected sequence motif are not highly dependent on either 2′-hydroxyl or 2′-amino moieties, then the same motif might be expected to be selected regardless of the chemical nature of the starting population. In this respect, it should be noted that the anti-elastase aptamers selected from a modified RNA pool have some residual binding activity when the sequence is composed solely of ribotides.[77] Thus, while it is likely that DNA and RNA of the same sequence assume quite different structures, the RNA and modified RNA of the same sequence may be more structurally similar.

B. Conformationally Restraining Minimal Oligonucleotides

Although modified aptamers can resist degradation, their size (generally greater than 30 nucleotides in length) may remain a barrier to cost-effective synthesis. Individual residues in an aptamer either directly contact small molecule or protein targets or else provide a structural context for those residues that do make direct contact. An easy way to observe this is to examine how the functional residues and secondary structures were determined for an anti-RT aptamer.[83] The original selection produced a series of ligands which contained a conserved se-

quence motif; this motif could be folded into a pseudoknot structure (**20**). The motif was synthe-



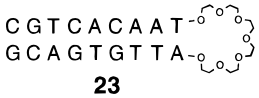
20

sized as a "doped" sequence pool and functional variants were selected. Sequence analysis of the selected species demonstrated that some residues could not vary without loss of binding function; these presumably directly contacted the protein or precisely positioned other chemical moieties to contact the protein. Other residues covaried in a manner that was consistent with the formation of a Watson−Crick paired stem. While the identity of these residues was relatively unimportant for function, the structural scaffold they formed was important for function. Overall, the selection experiments proved that a short (27-residue) RNA pseudoknot was sufficient for high-affinity interactions with HIV-1 RT.

Chemical linkers can be used to reduce the size of scaffolding elements (and the overall cost of synthesis) while retaining the positions of functional residues and moieties. A range of chemical and enzymatic methods are available to incorporate interstrand cross-links into the helical regions of DNA and RNA.[84] Alternatively, nucleotide loops have been used to "cap" functional nucleic acids at one (hairpin, **21**) or both (double-hairpin or "dumbbell", **22**) ends of a helix (Figure 6).[85−88] Certain loop sequences, such as GCGAAGC in DNA[89] and UUCG in RNA,[90]

21

22

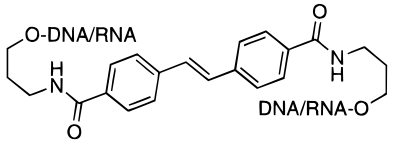


23

24

Figure 6. Schema of hairpin and dumbbell structures containing either nucleotide or non-nucleotide linkers. Double-hairpins can be made synthesizing an oligonucleotide in which the termini are aligned by a complementary strand and abut one another, enzymatically phosphorylating the oligonucleotide with T4 polynucleotide kinase, and ligating with DNA or RNA ligase. Alternatively, dumbbells can be made chemically by phosphorylating during solid-phase synthesis and ligating using either 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC)[87] or CNBr.[88] Yields are typically in the 30−50% range, however, there have been examples of yields as high as 70%.

have been found to be particularly effective at stabilizing structure and imparting nuclease resistance. Non-nucleotide linkers, such as (poly)ethylene glycol (PEG), have also been used to cap helices at one (**23**) or both (**24**) ends of the helix.[91] These reagents are easier and cheaper to make, and are commercially available. Structures constrained by PEG caps have been used to probe nucleic acid: protein interactions.[92] In addition, Kool and co-workers have shown that PEG linkers can increase the stability of oligonucleotides toward exonucleases.[93] Similarly, Lestinger has shown that a stilbenedicarboxamide linker (**25**) can increase the stability of an oligonucleotide.[94] The stilbene linker is rigid



25

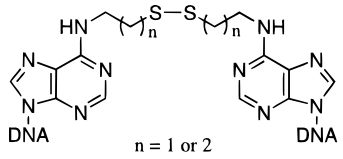
enough to effectively constrain the terminus of the helix, but also stacks with and stabilizes the helix.

These methods, however, have some potentially serious drawbacks. First, in some cases the linkers can distort or alter native helical geometry. Second, construction of dumbbell oligonucleotides can be technically difficult and in many cases poor overall yields are obtained. Finally, chemical linkers frequently span the terminal 3′- and 5′-hydroxyls of an oligonucleotide and interfere with enzymatic radiolabeling of the sample for biochemical assays such as footprinting and sequencing.

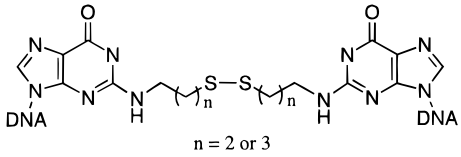
With knowledge of these drawbacks, a more recent technique to constrain oligonucleotides has emerged. Reactive functional groups have been engineered into oligonucleotides so that they are in close proximity with one another in the final structure and thus specifically react to form cross-links. Of these, thiols have been the reactive functional group of choice because: (1) disulfide bonds are formed in high yield, often quantitatively;[95] (2) the mild redox chemistry of thiol−disulfide interconversion provides little potential for undesired side reactions;[96] (3) disulfides are stable to a wide variety of solvents and reagents;[97,98] and (4) engineering specific conformations using disulfide cross-links is well-precedented in the peptide and protein literature.[99,100]

Lipsett first reported the formation of disulfide cross-links in a nucleic acid in 1966.[101−103] In those experiments, she found that $tRNA^{Tyr}$ from *E. coli* that naturally contained 4-thiouridine residues could be oxidized with iodine to form an intramolecular disulfide cross-link. Unfortunately, the potential utility of disulfide cross-links for studying nucleic acid structure was not realized because other, critical analytical methods such as RNA sequencing were not available. Therefore, it was not until 1991 that Glick and Verdine independently reported that disulfide cross-links could be site-specifically introduced into oligonucleotides via solid-phase chemical synthesis.

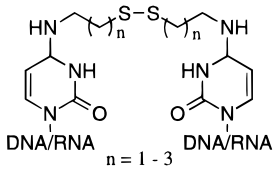
Verdine and co-workers have explored methods for introducing disulfide cross-links into the helical regions of nucleic acids.[104,105] In their method, thio-



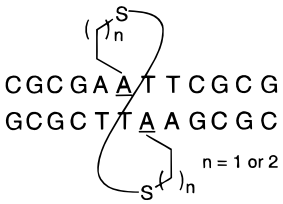
26



27



28

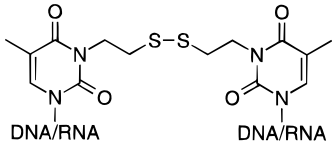


29

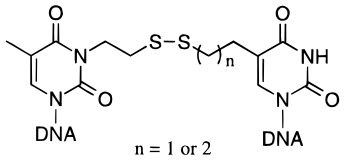
Figure 7. Thiol-modified nucleosides used by Verdine and co-workers to conformationally constrain oligonucleotides with disulfide cross-links (**26**−**28**).[104] These disulfide cross-links are located within the grooves of the helix and can be introduced at consecutive base pairs. The oligonucleotide shown in **29** is an example of how a disulfide can be introduced across the major groove of the helix.

alkyl tethers of varying lengths (3- to 5-atoms) are positioned at the N^6 of adenosine (**26**),[105] the N^2 of guanosine (**27**),[106] or the N^4 of cytosine (**28**).[107,108] The resultant disulfides occupy the grooves of the helix (Figure 7). In their first report, thiol-modified adenosines were placed at consecutive base pairs on opposite strands of a duplex (**29**) and under oxidative conditions formed an interstrand disulfide cross-link in the major groove of the helix. These cross-links imparted increased thermal stability ($\Delta T_m = 19$ °C) to the duplex relative to the unmodified sequence and resulted in only minimal distortion of its native geometry. Since this initial report, Verdine and co-workers have used disulfide chemistry to trap four different oligonucleotide conformations: a bent duplex,[109] a minor groove cross-link,[106] a torsionally strained duplex,[107] and a RNA hairpin.[108]

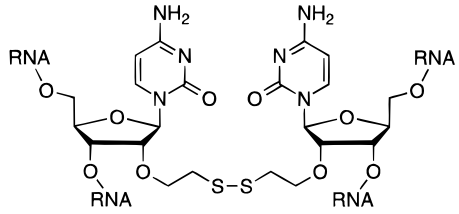
An alternative means to stabilize nucleic acids via disulfide cross-links has been described by Glick and co-workers.[110,111] In their work, alkyl mercaptan groups are appended to bases (**30** and **31**) or sugars (**32**) to form disulfide cross-links at specific sites (Figure 8). In contrast to positioning disulfide bonds



30



31



32

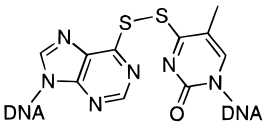
33

Figure 8. Thiol-modified nucleosides used by Glick and co-workers to conformationally constrain oligonucleotides with disulfide cross-links (**30**−**32**).[111] Disulfide cross-link formation is designed to occupy sterically accessible spaces and therefore to be structurally nonperturbing. For example, a duplex can be stabilized (**33**) by placing N^3-(thioethyl)thymidines (**30**) at the termini so that the cross-link resides at the ends of the helix.[117]

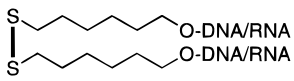
in the grooves of helices, these cross-links occupy sterically accessible spaces and therefore should conformationally stabilize nucleic acids without significant perturbation of their native structures. Using this chemistry, disulfide bonds have been introduced into nucleic acid secondary and tertiary structures such as hairpins[110,112] and duplexes,[111] triple helices,[113] and tRNA.[114] The structural[115,116] and thermodynamic[117,118] properties of these cross-linked oligonucleotides have been fully described. For example, when the terminal bases of the DNA dodecamer d(CGCGAATTCGCG)$_2$ (**33**) were replaced with N^3-(thioethyl)thymidines (**29**), the strands were prevented from dissociating and the thermal stability of the duplex was dramatically increased ($\Delta T_m = 38$ °C).[111] Thermodynamic analysis (UV and DSC) reveals that the enhanced stability is entropic in origin and structural analysis (NMR and CD) indicates that the helical structure has not been compromised.[117] In addition, preliminary evidence suggests that the cross-link prevents nucleolytic cleavage of the oligonucleotide using snake venom phosphodiesterase.[117] A final advantage of this method is that the cross-linked helices are compatible with enzymatic radiolabeling techniques.

Over the past several years, several additional methods for incorporating disulfide cross-links into DNA and RNA have been developed. The sulfur-bearing nucleotides 6-thioinosine and 4-thiothymi-

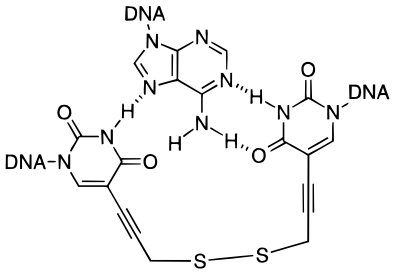
dine were introduced at the penultimate positions of a 12 base pair duplex to form an intermolecular cross-link (**34**).[119] Although the desired cross-linked product was achieved in high yield, the reaction was sluggish (five days to completion) and the disulfide bonds were shown to be quite unstable to the small amounts of reducing agents present during standard end-labeling procedures. In a procedure that is reminiscent of the introduction of PEG and stilbene linkers, Gao and co-workers have bridged the 5′- and 3′-hydroxyls of helices with disulfide loops (**35**).[120,121] A disulfide cross-linked circular oligonucleotide has also been reported.[122] In this work, two C^5-(thiopropyne)thymidines (**36**) were incorporated within a single oligonucleotide strand and upon binding a complementary stand in the major groove a cross-link results. The disulfide-linked circle binds to its single-stranded host with an association constant of about 10^{17} M^{-1}: 3 orders of magnitude greater than the biotin:strepavidin couple. Lastly, Eckstein and co-workers have developed a method for incorporating disulfide cross-links into a hammerhead ribozyme.[123] Specifically, thiols were introduced at the 2′-hydroxyls of specific pyrimidine bases (**37**) to probe the active conformation of the ribozyme.



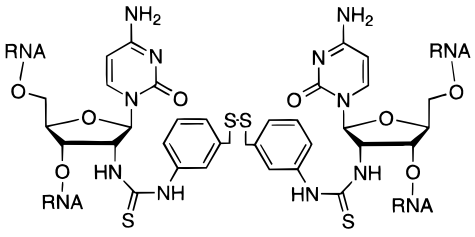
34



35

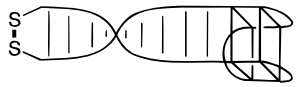


36

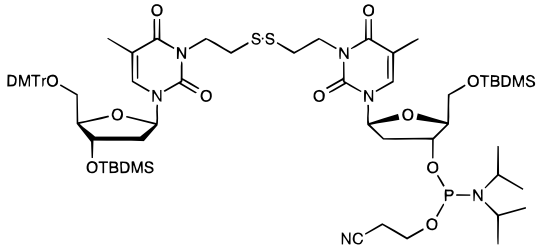


37

All of these methods are amenable to the stabilization of short aptamer sequences and structures. Nelson and co-workers have successfully introduced stilbene linkers at the termini of anti-Rev aptamers.[124] The cross-linked compounds bind Rev almost as well as the full-length *Rev responsive element* (*RRE*), which is over six times larger. In another example, Gao and co-workers linked the terminus of an anti-thrombin aptamer with either a PEG or disulfide cross-link (**35**) to give a the conformationally restrained G-tetramer (**38**).[74] Finally, Osborne and Ellington have developed a phosphoramidite (**39**) that can be used to introduce a disulfide cross-link at the termini of a helix during solid phase synthesis.[125] This reagent will be useful for constraining oligonucleotides that have difficulty forming properly and should pave the way for the synthesis of very short, very stable aptamers that are capped at both ends by disulfides.



38



39

VII. In Vivo Efficacy

In addition to blocking the activities of individual enzymes or regulatory proteins, aptamers can be used to inhibit particular aspects of cellular or viral metabolism in the same way that antisense oligonucleotides or triple helix agents are used to block genetic expression. For example, aptamers have been shown to block the function of a number of extracellular cytokines. When aptamers selected to bind to bFGF were assayed for their ability to inhibit bFGF binding to cell-surface receptors they were found to be effective at concentrations as low as 5 nM for low-affinity sites and 0.2 nM for high-affinity sites.[52] The anti-bFGF aptamers selected from modified RNA pools could inhibit interactions with receptors at concentrations as low as 1 nM for low-affinity sites and 3 nM for high-affinity sites.[78] An illustration of how anti-bFGF aptamers may function is given in Figure 9. Moreover, bFGF-induced cellular proliferation was blocked by aptamer concentrations of 50−100 nM. Similarly, anti-VegF aptamers selected from RNA pools inhibited receptor binding with an ED_{50} of 20−40 nM,[126] while the anti-VegF aptamers selected from modified RNA pools inhibited receptor binding at concentrations as low as 1 nM.[41] Finally, anti-IgE aptamers have been shown to block interaction with the Fce RI receptor and thereby inhibit IgE-mediated serotonin release from cells in tissue culture.[178]

Aptamer inhibition of enzymatic activity also has metabolic consequences. Aptamers selected against thrombin have been shown to block blood clotting in standard assays.[58] In addition, anti-thrombin aptamers can potentially prevent reocclusion of blood ves-

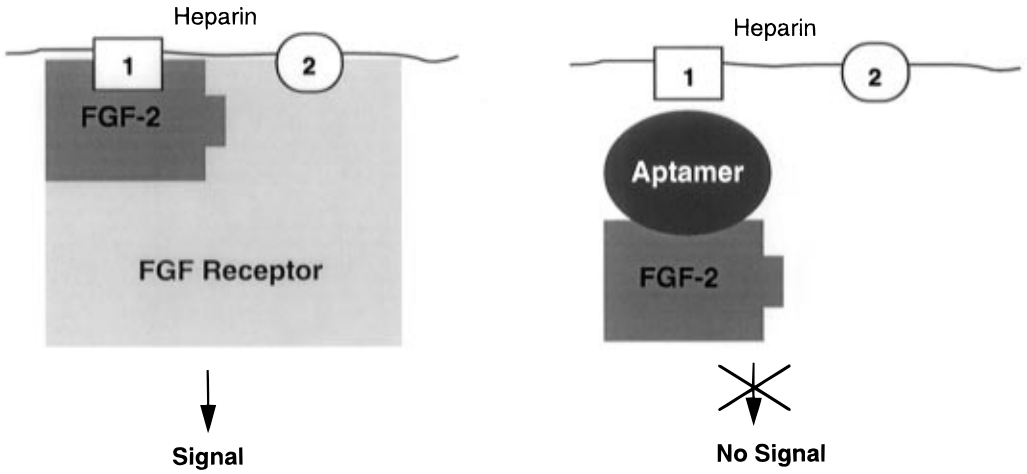


Figure 9. Illustration of aptamer inhibition of heparin binding sites. (Left) Receptor signaling requires formation of a trimeric complex where FGF-2 binds heparin at site 1 and to the FGF receptor (which binds the heparin at site 2). (Right) Aptamer binds to the heparin binding site of FGF-2 and prevents binding of the FGF receptor thereby suppressing the signaling event.

sels following coronary artery enzymatic thrombolysis or angioplasty.[127] In fact, in certain applications, such as hemodialysis or hemofiltration, the limited *in vivo* stability of the anti-thrombin aptamer may actually be advantageous, since clotting would be selectively inhibited for only a short period of time.[128] Anti-elastase aptamers that covalently link to their target were perfused into a rat lung and assayed for their ability to inhibit interleukin-1 (IL-1) induced, neutrophil-mediated damage (as monitored by edema-like weight gain).[72] The aptamer conjugated to the elastase inhibitor reduced the inflammatory response to near levels seen in the absence of IL-1 induction.

Finally, aptamers can modulate complex metabolic processes. Sequences selected from RNA and modified RNA populations to bind Rous sarcoma virus (RSV) can inhibit viral replication when mixed with the virus prior to infection.[81] An aptamer selected to bind a tRNA synthetase conjugated with its cognate amioadenylate was found to stimulate hydrolysis of noncognate aminoadenylates.[129] In other words, the aptamer enhanced the fidelity of error correction by an enzyme critical to protein biosynthesis. Anti-Rev aptamers have been shown to substitute for the Rev-binding element and efficiently facilitate RNA transport *in vivo*.[130]

In touting aptamers as potential pharmaceuticals we have obviously ignored a major problem: delivery. This problem has already been extensively considered (although not solved) in the development of antisense therapies, and many of the methods that have been developed for the delivery of antisense oligonucleotides are applicable to the delivery of aptamers. While assessments of delivery systems are outside the scope of this review, which focuses on selection principles and selected nucleic acids, we will present a brief rendition of the problems and prospects for aptamer delivery. It is unlikely that most nucleic acids will be orally available, although the possibility that some sequences or structures may directly enter the bloodstream should not be discounted. Therefore, aptamers against extracellular targets could be directly injected into an organism. Aptamers selected or engineered to resist degradation in sera can potentially be used "neat", without further protection. "Naked" or unprotected RNA or DNA aptamers may require further protection, most likely by conjugation to liposomes or other biopolymer carriers such as polyethylene glycol. Aptamers against intracellular targets can be delivered "endogenously" or "exogenously". Endogenous delivery is essentially gene therapy, with an anti-protein aptamer being transcribed in bulk from a strong polIII promoter. For example, anti-Rev aptamers expressed intracellularly have been shown to inhibit HIV-1 replication.[131] Exogenous delivery involves facilitated transport of aptamers across membranes. Polylysine has been successfully used to squelch the negative charges on nucleic acid backbones and enhance membrane transport.[132,133] A variety of cationic liposomal formulations have been developed for gene delivery, and these should be equally applicable to the delivery of aptamers.[134−136] Liposomes can even be targeted to specific cell types. For example, liposomes conjugated to folate can deliver nucleic acids to tumor cells.[137] Surprisingly, even unprotected nucleic acids may be delivered exogenously. Short oligonucleotides can be internalized by cellular pinocytosis[138] and endocytosis.[139] A variety of cell lines, including human cell lines such as HeLa, H9, and HL60, have been observed to take up exogenously introduced oligonucleotides rapidly and efficiently.[140−142] For example, antisense effects on intracellular targets can often be observed at external oligonucleotide concentrations ranging from 1−50 μM.[143] The internalized oligonucleotides typically localize to the nucleus.[144]

VIII. Conventional Drug Design

There are two ways in which *in vitro* selection can be used to further the development of conventional drugs (small organics). First, by helping to identify targets for drugs, either by defining nucleic acid targets or validating protein targets. Functional nucleic acids represent a largely untapped source of biological targets for drug development. Beyond their role as information-carrying macromolecules, there are numerous nucleic acids whose sequences and structures are crucial to viral, cellular, or organismal

function. Microorganisms have already discovered this fact and have evolved secondary metabolites, such as aminoglycoside antibiotics, that can bind to and inhibit cellular RNA molecules, such as ribosomal RNA. Aminoglycosides have also been shown to fortuitously inhibit other functional nucleic acids, including the Rev responsive element of HIV-1, the group I self-splicing intron, and the hammerhead ribozyme. *In vitro* selection can be used to define the binding domains of potential nucleic acid targets. In addition, the exquisite specificities of selected nucleic acids makes them useful tools for dissecting metabolism and validating whether other targets, such as proteins, are worthy candidates for drug development. Second, *in vitro* selection can be used to identify drugs themselves. Selection can assist in identifying putative anti-nucleic acid drugs, especially those derived from combinatorial chemical libraries.

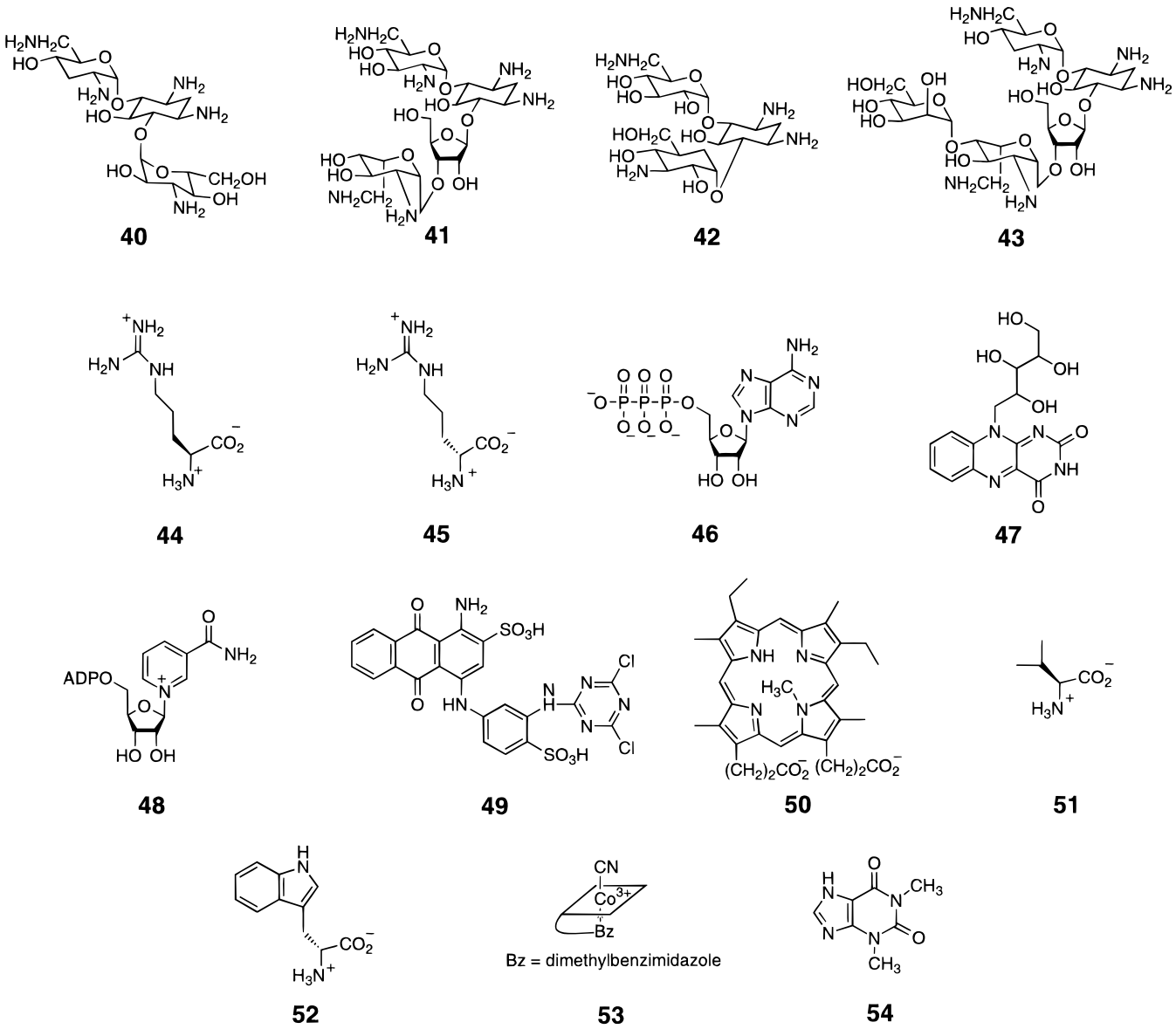
A. Target Identification

The sequences and structures of nucleic acid targets can be precisely defined by *in vitro* selection. As was the case with the anti-HIV aptamer described above, by sieving binding from nonbinding variants from a partially randomized population of sequences the relative importance of individual residues can be quickly established. Bartel and co-workers utilized this procedure to define the *Rev responsive element* of HIV-1.[145] The *RRE* had previously been partially defined by deletion analysis; a 67 nucleotide tract known to be important for function was synthesized so that individual positions contained 65% wild-type residues, 30% non-wild-type residues (10% of each non-wild-type residue), and 5% deletions. This population contained all single through hextuple variants of the 66 nucleotide sequence. Variants that could bind to Rev were selected by filtration. After three cycles of selection and amplification the population could bind better than the original wild-type *RRE*. The variants were sequenced, and 20 residues presented within a short stem-internal loop-stem structure were found to be important for Rev-binding activity. This delimited RNA is a potential target for drug development. In fact, the aminoglycoside antibiotic neomycin has been shown to fortuitously disrupt interactions between Rev and the *RRE*.[146] The neomycin binding site has been localized to the stem-internal loop-stem structure by modification interference analysis, and the drug was found to contact several of the residues revealed by selection experiments to be important for function. Baskerville and co-workers have carried out similar studies with the Rex-binding element of HTLV-I and have determined that residues within a stem containing two base bulge loops are important for Rex-binding activity.[147] Completely random sequence pools can also be used to define functional RNA sequences, providing that the random sequence tracts are short. For example, two hexanucleotide loops in the hepatitis B virus encapsidation signal were separately randomized, and functional variants were selected *in vivo* based on their ability to be encapsidated.[148] Sequence comparison of selected clones revealed that most of the randomized positions had reverted to wild-type and hence were essential for RNA packaging. These results suggest that the virus might not readily mutate to avoid inhibition by a drug that targeted one of the loops. However, it should be noted that while selection strategies can provide insights into the sequences and structures of natural RNAs, they do not always return wild-type binding sites. When the neomycin binding site of ribosomal RNA was partially randomized and selected for its ability to interact with neomycin, the sequences that were recovered bore no resemblance to the wild-type.[149] Instead, a motif previously observed in anti-neomycin aptamers selected from completely random sequence pools was recovered.[150]

Selection can also be used to define DNA targets. Because DNA binding sites for proteins are located within a similar structural context (the double helix), *in vitro* selection experiments typically start with a completely random pool of double-stranded DNA. Binding sites for several members of the basic helix-loop-helix family of transcription factors, including MyoD, have been defined by this method.[25] Interestingly, it looks as though the recognition sequences for homo- and heterodimeric transcription factors may be modular, with each protein monomer recognizing its particular half-site. These results have recently been extended to several other members of the bHLH family, including the Ah receptor, the Ah receptor nuclear transport protein, and single-minded protein.[151] These results are especially significant given the recent development of "lexitropsins" (sequence-reading compounds) that can specifically recognize stretches of DNA that are similar in length to the bHLH family binding sites.[152,153]

Selection can help to define targets other than nucleic acids. As we have seen, aptamers can frequently distinguish between closely related protein targets. It may therefore be possible to use selected nucleic acid shapes as probes that assess whether or not drugs that bind a particular site are likely to be cross-reactive. *In vitro* selection experiments may even be able to uncover binding specificities that were previously unknown. Experiments with polyprimidine tract-binding proteins U2AF,[65] Sex-lethal, and polyprimidine tract-binding protein (PTB) provide a particularly good example of how selections can reveal new specificities.[154] All of these proteins bind to predominantly uridine-rich sequences. Uridine-rich sequences selected by U2AF[65] were similar to polyprimidine tracts found in a variety of genes, consistent with the role of U2AF[65] as a general splicing factor. In contrast, Sex-lethal yielded a specific uridine-rich sequence similar to the pyrimidine tract of substrates such as pre-mRNA. Selection against PTB yielded a distinct consensus sequence found within the introns of rat α- and β-tropomyosin. Similarly, selections that targeted the ASF/SF2 (alternative splicing factor/splicing factor 2) and SC35 (splicing component 35) splicing factors resulted in the isolation of distinct consensus-sequence motifs, indicating that these factors may not be functionally redundant.[155] Aptamers can not only probe the specificities of targets related by history or function, but can also be used to probe different states or conformations of the same target. For example,

Chart 2. Structure of Small Molecule Ligands Used as Aptamer "Guests"



selections that targeted whole ribosomes identified different aptamers depending on whether ribosomal protein S1 was present or absent from the complex.[156]

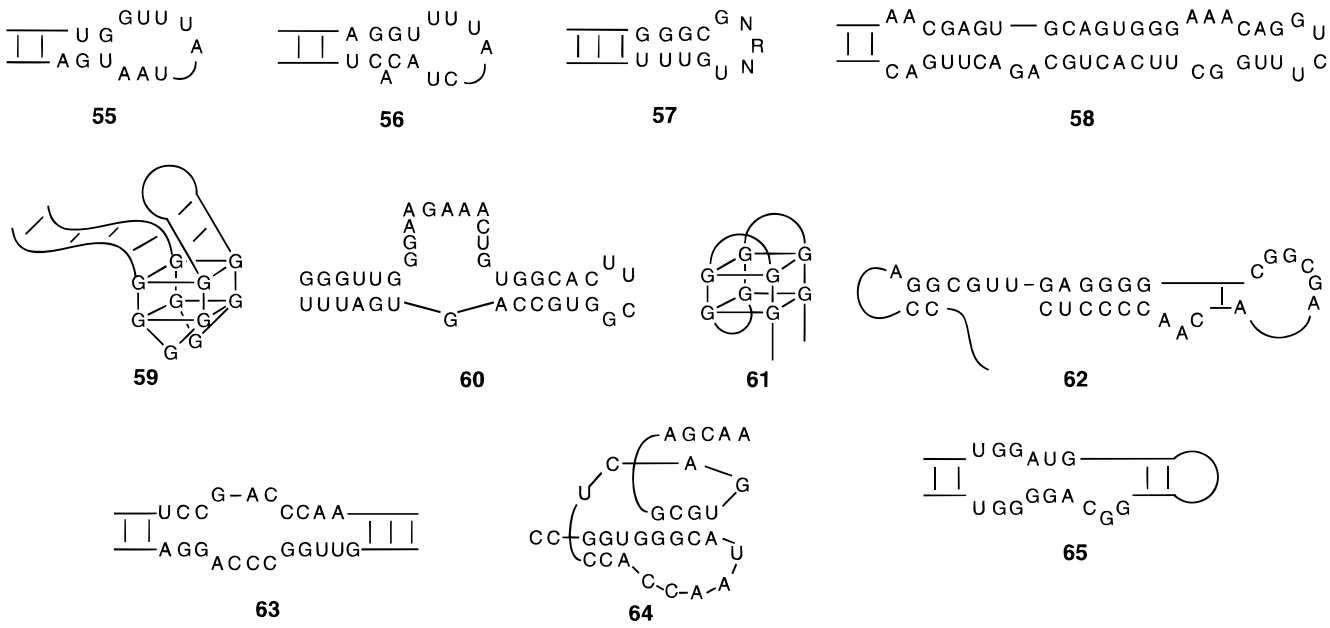
While splicing factors and the ribosome were already known to bind nucleic acids, aptamers could similarly be used to probe the specificities of non-nucleic acid binding protein targets. For example, while many cytokines and receptors possess what are nominally known as heparin binding sites, the architectures of these sites and their natural oligosaccharide ligands may be quite different. Anti-FGF aptamers compete with heparin for binding to FGF,[52,78] anti-VegF aptamers compete with heparin for binding to VegF,[126] and anti-thrombin aptamers selected from RNA pools compete with heparin for binding to thrombin,[157] but these nucleic acid ligands are not known to cross-recognize their protein targets. The structural and functional differences between these aptamers may therefore signal differences in oligosaccharide binding site specificities. To the extent that aptamers have *in vivo* efficacy, such binding differences can potentially be translated into differential activities in biological assays. In this way, aptamers can be used to validate a protein target by determining whether inhibition of a given protein will produce a given physiological effect. In effect, even when aptamers may not be suitable drugs in and of themselves, they can be used as drug mimics.

B. Drug Design

In vitro selection can also be used to discover drugs rather than define targets. The discovery process can be applied either to existing drugs or to novel compounds that bind nucleic acids. Nucleic acid sequences that are discovered to bind to existing drugs may have counterparts in organismal or viral genomes, and hence may reveal new targets for compounds with proven bioavailabilities and pharmacokinetic properties.

While we have so far focused primarily on large biomolecular targets, aptamers that specifically bind to small molecular "guests" can also be selected (Chart 2). For example, the aminoglycoside antibiotics tobramycin (**40**), neomycin (**41**), kanamycin (**42**), and lividomycin (**43**) have elicited tight and specific aptamers. Wang and Rando examined the binding interactions of tobramycin with an RNA pool con-

Chart 3. Consensus Sequences and Secondary Structures of Aptamers Selected To Bind Small Molecule Ligands



taining a 60-base random site.[158] In their study they found a number of aptamers that weakly bound (~10 mM) tobramycin with no apparent consensus sequence. However, upon increasing the stringency of selection by lowering the amount of ligand attached to a solid support, the affinity for binding increased with each cycle until the binding was in the submicromolar range. Cloning of the aptamers revealed two similar stem-loop consensus sequences (**55** and **56**) (Chart 3). Similarly, selections that targeted neomycin generated a consensus motif (**57**) that could interact tightly with the antibiotic ($K_D \approx 100$ nM) and discriminate against the aminoglycoside paromomycin (**66**), which differs by only a single amino moiety, by over 100-fold.[150] Finally, aptamers that can form complexes with kanamycin (**42**) and lividomycin (**43**) have been selected from a random sequence RNA pool.[159] After several cycles, the selected species had affinities (aggregate $K_D \approx 300$ nM) and specificities for the aminoglycosides that rivaled those of ribosomal RNA. As was the case for the early cycles of the selections that targeted tobramycin, there were numerous ($\approx 10^6$ estimated) aptamers in the selected population and no clear consensus sequence. The fact that multiple, different sequences exist that have the same binding characteristics as a known drug target (ribosomal RNA) implies that there may be many cellular or viral RNA molecules that could also be bound and affected by aminoglycosides. A map of aminoglycoside structure to RNA sequence might provide new targets for old drugs. In order to test this "reverse drug discovery" method, the lividomycin aptamers were further selected until a consensus sequence emerged (**58**).[160] The consensus sequence was compared with all sequences in Genbank, and a number of potential targets, including RNAs from *Haemophilus influenzae* and the parasite *Leishmania*, were identified. Because small molecules recognize relatively small motifs (10−30 bases), there is a good chance that at least one gene or structural motif will be discovered.

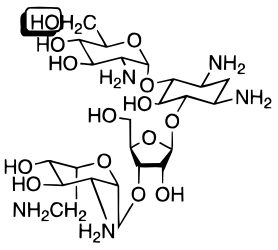
There is reason to believe that there may be a large number of drug leads beyond aminoglycosides that might be candidates for such reverse drug discovery methods. Some of the earliest work in understanding the applications of selection techniques was performed with small organic dyes (**49**) which previously had not been known to interact with nucleic acids.[29,161] Natural as well as synthetic compounds have also been shown to have surprising affinities for particular RNA molecules. Lauhon and Szostak targeted the biological cofactor, nicotinamide adenine dinucleotide (NAD) (**48**).[162] Selection of an RNA pool containing an 80-base random region against NAD using nicotinamide mononucleotide (NMN) (**67**) as an eluant resulted in nearly 50% of the population being selectively bound after eight cycles of selection. One of the clones isolated from the final cycle could discriminate between the oxidized cofactor, NAD^+ (K_D of 2.5 mM), and the reduced cofactor, NADH (K_D of 37 mM), which differ by only a single hydride. The selectivity exhibited by this RNA is similar to that seen for protein dehydrogenases and reductases. Two further examples of aptamers that bind small organic molecules with particularly high affinities and specificities were afforded by selections that targeted ATP (**46**) and the drug theophylline (**54**). The anti-ATP aptamer (**59**) selected by Sassanfar and Szostak has a K_D for ATP of 0.7 μM and discriminates against a closely related substrate, dATP, by a factor of over 1000.[163] The anti-theophylline aptamer (**64**) selected by Jenison and co-workers has a K_D for theophylline of 320 nM.[164] During the selection procedure, aptamers that could also recognize caffeine (**68**), which differs from theophylline by the addition of a single methyl group, were removed from the population. As a result of this counterselection, the antitheophylline aptamer can discriminate against caffeine by a factor of over 10 000, a value better than that observed for comparable monoclonal antibodies. A variety of other small molecule targets, from compounds as small as zinc[165,166] to compounds as large as porphyrins (**50**)[167]

Table 5. Aptamers Selected To Bind Small Molecule Ligands

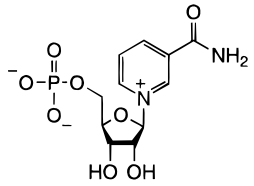
ligand	entry number[a]	pool	selected aptamer[b]	conjugate ligand K_D (μM)	specificity (nonconjugate)	ref
tobramycin	**40**	RNA	**55** and **56**	0.006	900 (6′-*N*-4-pyreneacetyl tobramycin)	158
neomycin	**41**	RNA	**57**	0.10	100 (paromomycin)	150
kanamycin	**42**	RNA	N/A[c]	≈0.3	10−100 (kanamycin B)	159
lividomycin	**43**	RNA	**58**	≈0.3	≈1 (paromomycin)	159
L-arginine	**44**	RNA	**6**	10.0	120 (citrulline)	50
D-arginine	**45**	RNA	**3**	200	1.7 (L-arginine)	47
ATP	**46**	DNA	**59**	6	N/A	82
	46	RNA	**60**	0.7	1430 (dATP)	163
riboflavin	**47**	RNA	**61**	1−5	~1 (5-deazariboflavine)	162
nicotinamide (NAD)	**48**	RNA	**62**	2.5	14.8 (NADH)	162
reactive blue 4	**49**	RNA	**63**	600	85 (cibacron blue)	29
N-methylmesoporphyrin IX	**50**	DNA	N/A	0.5	3 (mesoporphyrin IX)	167
L-valine	**51**	RNA	N/A	2900	62 (D-valine or leucine)	*d*
D-tryptophan	**52**	RNA	N/A	18.0	670 (L-tryptophan)	*e*
vitamin B12 (cyanocobalamin)	**53**	RNA	**64**	0.088	11400 (adenosyl-cobalamin)	168
theophylline	**54**	RNA	**65**	0.32	10900 (caffeine)	164

[a] See Chart 2 for ligand structures. [b] See Chart 3 for aptamer structures. [c] Not available or determined. [d] Majerfeld, I.; Yarus, M. *Nature Struct. Biol.* **1994**, *1*, 282. [e] Famulok, M.; Szostak, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 3990.

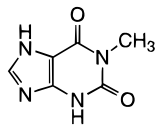
and vitamin B12 (cyanocobalamin, **53**),[168] have successfully elicited aptamers. Selections that targeted small molecules was recently reviewed by Ellington and Table 5 provides an updated summary of results.[9]



66

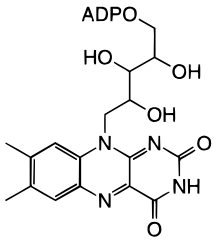


67

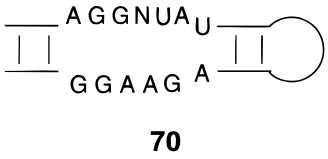


68

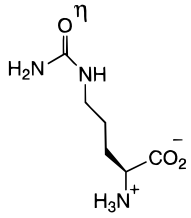
Not only are the binding characteristics of selected nucleic acids consonant with the use of small molecules as anti-nucleic acid drugs, but the recently determined structures of aptamer:ligand complexes have features that are reminiscent of protein:drug complexes. Burgstaller and Famulok selected aptamers that could form complexes with the cofactor FMN (**69**), and used comparative sequence analysis to derive a minimal, functional motif (**70**).[169] Fan and co-workers have determined the NMR structure of the complex.[170] Recognition is based on stacking interactions below and above the aromatic, planar cofactor; for example, a base triple (A:U:G) lies parallel and adjacent to the bound FMN. Some specificity for FMN is provided by two hydrogen bonds between the cofactor and the Hoogsteen face of an adenosine that lies in the same plane. These features are similar to those observed in complexes between trimethoprim and dihydrofolate reductase. Famulok also selected aptamers that could bind to either citrulline (**71**) or arginine (**44**), depending on the identity of three positions in the selected motif.[50]



69



70

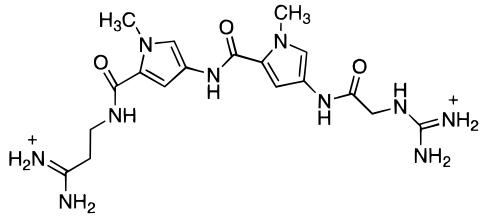


71

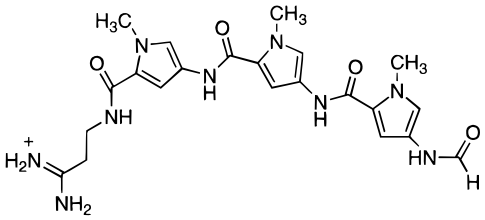
Modification interference experiments carried out with a minimal binding motif revealed that an internal loop structure was critical for interactions with the amino acids, and that the conformation of the aptamer changed on complexation.[171] The NMR structures of the anti-citrulline and anti-arginine aptamers complexed with their respective ligands

have been determined.[172] Five guanosine residues come together to form the base of a binding pocket; the aliphatic side chains of the amino acids stacks over this base. The binding pockets are lined with hydrogen-bond donors and acceptors that contact the side chains of the amino acids. In the anti-citrulline aptamer, O^{η} is contacted by the hydrogen at the N^3 position of U13 and the exocyclic amine of G29; in the anti-arginine aptamer, the corresponding N^{η} position is contacted instead by the N^3 of C13: a hydrogen-bond donor has been exchanged for a hydrogen-bond acceptor. Surprisingly, there are no apparently electrostatic contacts between the aptamer and arginine, and the interactions are unlike those seen between nucleic acids and the arginines or nucleic acid binding proteins. The structure of the anti-ATP aptamer has also been solved by NMR.[173] As was the case with the anti-amino acid aptamers, the ligand is bound in a loop between two stems. However, the overall architecture of the complexes are quite different. The adenine base is intercalated between two purines, A10 and G11; G11 pinches off a short loop structure by forming a non-Watson−Crick interaction with G7. This loop bears a striking resemblance to a well-known and previously characterized RNA structure, the GNRA tetraloop, with the adenosine ligand base-pairing with G8 and acting as the 3′ most adenosine of the loop. The right-hand stem of the structure is also capped by a non-Watson−Crick interaction (between G17 and G34), and provides additional hydrogen bonds to the N^3 position of adenine and to the free hydroxyls of ribose. The pocket that is formed contacts nearly half of the available surface area of the ligand, and satisfies almost every possible hydrogen bonding partner.

The selectivities and structures of aptamers that bind small molecules lend credence to the suggestion that *in vitro* selection can help to form the basis of a drug development program.[174] While there are numerous compounds that have been designed to bind nucleic acids, their known affinities and specificities are too low to ensure their success as drugs. For example, netropsin (**72**) and distamycin (**73**) have



72



73

served as starting points for attempts for the design and construction of "lexitropsins", but these compounds and their derivatives generally have dissociation constants from nucleic acid targets (A:T tracts) in the micromolar range[175] and discriminate against noncognate sequences by factors of only 10−100.[176] The aminoglycosides in general have slightly higher affinities for their targets, but are still far too catholic in their preferences for nucleic acids. However, recent attempts to augment the number of residues recognized by the minor-groove binding lexitropsins and to use aminoglycosides as synthetic scaffolds for the production of novel pharmacophores are promising. These efforts can be facilitated by an examination of the problem from the point of view of the biopolymer "host," rather than the organic "guest". *In vitro* selection can define and rank the sequence motifs that are recognized by anti-nucleic acid drug leads, such as lexitropsins or aminoglycosides. As new synthetic derivatives are produced, their specificities can be determined, and the process iterated to hone interactions between a drug and its desired target. A database that maps compound structure to nucleic acid sequence should eventually prove invaluable in the *de novo* design of specific anti-nucleic acid compounds.

The fact that a plethora of compounds have successfully elicited aptamers immediately suggests that many excellent anti-nucleic acid drug leads are currently unknown. *In vitro* selection could therefore be routinely used to screen existing chemical libraries, providing information as to which compounds bind which nucleic acid sequences or structures. Given the defined nature of the selection process, it is likely that the entire procedure, including sequence acquisition, could be automated. More importantly, nucleic acid selections could be carried out against mixtures of compounds. This should allow wholescale screens of chemical libraries by random sequence nucleic acid pools: the compounds with the highest affinity for nucleic acids in general would elicit binding species; these binding species would compete with one another until only the best pairings between potential anti-nucleic acid drugs and aptamers remained. Further definition of the selected sequences by chemical probes or structural studies may then suggest rational chemical modifications that would increase the affinities of anti-nucleic acid drug leads. Screens of combinatorial chemistry libraries could be combined with the "reverse drug discovery" method to yield fortuitous "hits" in cellular or viral genomes.

C. Mimetics

Finally, aptamers themselves may serve as scaffolds for the design and synthesis of small organics. Just as peptidomimetics have been synthesized on the basis of peptides selected from phage display libraries or protein structures, so too can nucleic acid mimetics be synthesized on the basis of the sequences and structures of aptamers. Understanding how functional groups on the bases, sugars, and phosphates of aptamers are positioned in a protein active site will obviously be aided by a determination of either the aptamer structure or, better, the structure of the complex. If these structures are known, then the functional groups can be introduced in a similar

configuration on a small organic backbone. Unfortunately, as is the case with proteins, deriving the structures of nucleic acids and nucleic acid complexes by NMR or crystallography is a somewhat laborious task. However, unlike with proteins, it may be possible to readily model the structures of nucleic acids. While the overall form of a protein is largely a consequence of tertiary structural interactions, the overall form of a nucleic acid is largely a consequence of secondary structural interactions. Nucleic acid secondary structures are in turn primarily dependent on Watson−Crick and other base pairings that have relatively defined structures. Thus, once the secondary structure of a nucleic acid is known, and if some tertiary structural contacts are known or suspected, then a three-dimensional structural model can be readily generated. An example of this type of analysis has already been carried out with an aptamer that binds HIV-1 Rev.[177]

A more intriguing prospect for drug design involves using nucleic acids as scaffolds for the presentation of limited chemical libraries. The fact that selections can be carried out with modified nucleic acids immediately suggests that the modifications could be geared toward applications other than stability. For example, in order to "fill" a hydrophobic pocket on a protein, aptamers could be selected from libraries derivatized with hydrophobic functional groups. The selected sequences could then be used to model or physically map how the hydrophobic functional groups were positioned, and mimetics designed. In this case, though, the mimetics would not rely as heavily on the nucleic acid backbone, but rather would attempt to string together the identities, positions, and orientations of the hydrophobic functional groups. In this way, nucleic acid libraries might effectively become replicating chemical libraries.

IX. Conclusion

In vitro selection experiments can provide both lead compounds for drug development and insights into methods for drug discovery. Aptamers generally bind to their targets with high affinities and specificities, and can be converted into smaller, more stable compounds by a variety of chemical modifications. *In vitro* selection can be thought of as an adjunct to combinatorial chemistry. Traversing functional "landscapes" with random sequence nucleic acids is roughly analogous to traversing the same "landscapes" with combinatorial chemical libraries, and some of the same principles may generally apply. Selection can be used to define novel nucleic acid targets for drug discovery or design, and to validate more popular protein targets. The interplay between combinatorial chemical libraries and selected nucleic acids may yield "codes" for recognition that will serve as the basis for rational drug design.

X. Acknowledgments

This research was supported by a Camille-Dreyfus Teacher−Scholar, a National Science Foundation Young Investigator Award, and by NIH grant AI-36083 (A. D. Ellington).

XI. References

(1) Janda, K. D. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10779.
(2) Lebl, M.; Krchnak, V.; Sepetov, N. F.; Seligmann, B.; Strop, P.; Felder, S.; Lam, K. S. *Biopolymers* **1995**, *37*, 177.
(3) Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *BioTechniques* **1992**, *13*, 901.
(4) Ecker, D. J.; Vickers, T. A.; Hanecak, R.; Driver, V.; Anderson, K. *Nucleic Acids Res.* **1993**, *21*, 1853.
(5) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. *Science* **1991**, *251*, 767.
(6) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5381.
(7) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10700.
(8) Chabala, J. C. *Curr. Opin. Biotechnol.* **1995**, *6*, 632.
(9) Ellington, A. D. *Curr. Biol.* **1994**, *4*, 427.
(10) Ellington, A. D.; Conrad, R. *Biotechnology Annual Review*; El-Gewely, M. R., Ed.; Elsevier: Amsterdam, 1995; Vol. 1, p 185.
(11) Gold, L. *J. Biol. Chem.* **1995**, *270*, 13581.
(12) Gold, L.; Polisky, B.; Uhlenbeck, O.; Yarus, M. *Annu. Rev. Biochem.* **1995**, *64*, 763.
(13) Conrad, R. C.; Giver, L.; Tian, Y.; Ellington, A. D. *Methods Enzymol.* **1996**, *267*, 336.
(14) Fitzwater, T; Polisky, B *Methods Enzymol.* **1996**, *267*, 275.
(15) Uphoff, K. W.; Bell, S. D.; Ellington, A. D. *Curr. Opin. Struct. Biol.* **1996**, *6*, 281.
(16) Mills, D. R.; Peterson, R. L.; Spiegelman, S. *Proc. Natl. Acad. Sci. USA* **1967**, *58*, 217.
(17) Spiegelman, S. *Q. Rev. Biophys.* **1971,** *4,* 213.
(18) Kramer, F. R.; Mills, D. R.; Cole, P. E.; Nishihara, T.; Spiegelman, S. *J. Mol. Biol.* **1974**, *89*, 719.
(19) Orgel, L. E. *Proc. R. Soc. London (Biol.)* **1979**, *205*, 435.
(20) Oliphant, A. R.; Struhl, K. *Methods Enzymol.* **1987**, *155*, 568.
(21) Oliphant, A. R.; Struhl, K. *Nucleic Acids Res.* **1988**, *16*, 7673.
(22) Horwitz, M. S.; Loeb, L. A. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 7405.
(23) Horwitz, M. S.; Loeb, L. A. *J. Biol. Chem.* **1988**, *263*, 14724.
(24) Oliphant, A. R.; Brandl, C. J.; Struhl, K. *Mol. Cell. Biol.* **1989**, *9*, 2944.
(25) Blackwell, T. K.; Weintraub, H. *Science* **1990**, *250*, 1104.
(26) Pollock, R.; Treisman, R. *Nucleic Acids Res.* **1990**, *18*, 6197.
(27) Thiesen, H.-J.; Bach, C. *Nucleic Acids Res.* **1990**, *18*, 3203.
(28) Sun, X.-H.; Baltimore, D. *Cell* **1991**, *64*, 459.
(29) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
(30) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
(31) Green, R.; Ellington, A. D.; Bartel, D. P.; Szostak, J. W. *Methods* **1991**, *2*, 75.
(32) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411.
(33) Lorsch, J. R.; Szostak, J. W. *Nature* **1994**, *371*, 31.
(34) Ciesiolka, J.; Illangasekare, M.; Majerfeld, I.; Nickles, T.; Welch, M.; Yarus, M.; Zinnen, S. *Methods Enzymol.* **1996**, *267*, 315.
(35) Guatelli, J. C.; Whitfield, K. M.; Kwoh, D. Y.; Barringer, K. J.; Richman, D. D.; Gingeras, T. R. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 1874.
(36) Davis, J. P.; Janjic, N.; Javornik, B. E.; Zichi, D. A. *Methods Enzymol.* **1996**, *267*, 302.
(37) Bunin, B. A.; Plunkett, M. J.; Ellman, J. A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4708.
(38) Giver, L.; Bartel, D.; Zapp, M.; Pawul, A.; Green, M.; Ellington, A. D. *Nucleic Acids Res.* **1993**, *21*, 5509.
(39) Tuerk, C.; MacDougal-Waugh, S. *Gene* **1993**, *137*, 33.
(40) Jensen, K. B.; Green, L.; MacDougal-Waugh, S.; Tuerk, C. *J. Mol. Biol.* **1994**, *235*, 237.
(41) Green, L. S.; Jellinek, D.; Bell, C.; Beebe, L. A.; Feistner, B. D.; Gill, S. C.; Jucker, F. M.; Janjic, N. *Chem. Biol.* **1995**, *2*, 683.
(42) Hamm, J. *Nucleic Acids Res.* **1996**, *24*, 2220.
(43) Schuster, P.; Fontana, W.; Stadler, P. F.; Hofacker, I. L. *Proc. R. Soc. London Ser. B* **1994**, *255*, 279.
(44) Binkley, J.; Allen, P.; Brown, D. M.; Green, L.; Tuerk, C.; Gold, L. *Nucleic Acids Res.* **1995**, *23*, 3198.
(45) Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K.; Spellmeyer, D. C.; Tan, R.; Frankel, A. D.; Santi, D. V.; Cohen, F. E.; Bartlett, P. A. *Proc. Natl. Acad. Sci.* USA **1992**, *89*, 9367.
(46) Ellington, A. D. *Ber. Bunsen-Ges. Phys. Chem.* **1994**, *98*, 1115.
(47) Connell, G. J.; Illangesekare, M.; Yarus, M. *Biochemistry* **1993**, *32*, 5497.
(48) Connell, G. J.; Yarus, M. *Science* **1994**, *264*, 1137.
(49) Tao, J.; Frankel, A. D. *Biochemistry* **1996**, *35*, 2229.
(50) Famulok, M. *J. Am. Chem. Soc.* **1994**, *116*, 1698.
(51) Geiger, A.; Burgstaller, P.; von der Eltz, H.; Roeder, A.; Famulok, M. *Nucleic Acids Res.* **1996**, *24*, 1029.
(52) Jellinek, D.; Lynott, C. K.; Rifkin, D. B.; Janjic, N. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11227.
(53) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. *Science* **1995**, *269*, 364.
(54) Ekland, E. H.; Bartel, D. P. *Nucleic Acids Res.* **1995**, *23*, 3231.
(55) Chen, H.; McBroom, D. G.; Zhu, Y.-Q.; Gold, L.; North, T. W. *Biochemistry* **1996**, *35*, 6923.
(56) Eaton, B. E.; Gold, L.; Zichi, D. A. *Chem. Biol.* **1995**, *2*, 633.

(57) Conrad, R.; Keranen, L. M.; Ellington, A. D.; Newton, A. C. *J. Biol. Chem.* **1994**, *269*, 32051.
(58) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564.
(59) Wu, Q.; Tsiang, M.; Sadler, J. E. *J. Biol. Chem.* **1992**, *267*, 24408.
(60) Tsiang, M.; Gibbs, C. S.; Griffin, L. C.; Dunn, K. E.; Leung, L. L. K. *J. Biol. Chem.* **1995**, *270*, 19370.
(61) Tian, Y.; Adya, N.; Wagner, S.; Giam, C.-Z.; Green, M. R.; Ellington, A. D. *RNA* **1995**, *1*, 317.
(62) Brown, D.; Gold, L. *Biochemistry* **1995**, *34*, 14765.
(63) Macaya, R. F.; Schultze, P.; Smith, F. W.; Roe, J. A.; Feigon, J. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3745.
(64) Wang, K. Y.; McCurdy, S.; Shea, R. G.; Swaminathan, S.; Bolton, P. H. *Biochemistry* **1993**, *32*, 1899.
(65) Wang, K. Y.; Krawczyk, S. H.; Bischofberger, N.; Swaminathan, S.; Bolton, P. H. *Biochemistry* **1993**, *32*, 11285.
(66) Schultze, P.; Macaya, R. F.; Feigon, J. *J. Mol. Biol.* **1994**, *235*, 1532.
(67) Paborsky, L. R.; McCurdy, S. N.; Griffin, L. C.; Toole, J. J.; Leung, L. L. K. *J. Biol. Chem.* **1993**, *268*, 20808.
(68) Padmanabhan, K.; Padmanabhan, K. P.; Ferrara, J. D.; Sadler, J. E.; Tulinsky, A. *J. Biol. Chem.* **1993**, *268*, 17651.
(69) Nieuwlandt, D.; Wecker, M.; Gold, L. *Biochemistry* **1995**, *34*, 5651.
(70) Xu, W.; Ellington, A. D. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7475.
(71) Lin, Y.; Padmapriya, A.; Morden, K. M.; Jayasena, S. D. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 11044.
(72) Smith, D.; Kirschenheuter, G. P.; Charlton, J.; Guidot, D. M.; Repine, J. E. *Chem. Biol.* **1995**, *2*, 741.
(73) Jensen, K. B.; Atkinson, B. L.; Willis, M. C.; Koch, T. H.; Gold, L. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 12220.
(74) Macaya, R. F.; Waldron, J. A.; Beutel, B. A.; Gao, H.; Joesten, M. E.; Yang, M.; Patel, R.; Bertelsen, A. H.; Cook, A. F. *Biochemistry* **1995**, *34*, 4478.
(75) Schneider, D. J.; Feigon, J.; Hostomsky, Z.; Gold, L. *Biochemistry* **1995**, *34*, 9599.
(76) Latham, J. A.; Johnson, R.; Toole, J. J. *Nucleic Acids Res.* **1994**, *22*, 2817.
(77) Lin, Y.; Qiu, Q.; Gill, S. C.; Jayasena, S. D. *Nucleic Acids Res.* **1994**, *22*, 5229.
(78) Jellinek, D.; Green, L. S.; Bell, C.; Lynott, C. K.; Gill, N.; Vargeese, C.; Kirschenheuter, G.; McGee, D. P. C.; Abesinghe, P.; Pieken, W. A.; Shapiro, R.; Rifkin, D. B.; Moscatelli, D.; Janjic, N. *Biochemistry* **1995**, *34*, 11363.
(79) Pieken, W. A.; Olsen, D. B.; Benseler, F.; Aurup, H.; Eckstein, F. *Science* **1991**, *253*, 314.
(80) Green, L.; Waugh, S.; Binkley, J. P.; Hostomska, Z.; Hostomsky, Z.; Tuerk, C. *J. Mol. Biol.* **1995**, *247*, 60.
(81) Pan, W.; Craven, R. C.; Qiu, Q.; Wilson, C. B.; Wills, J. W.; Golovine, S.; Wang, J.-F. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 11509.
(82) Huizenga, D. E.; Szostak, J. W. *Biochemistry* **1995**, *34*, 656.
(83) Tuerk, C.; MacDougal, S.; Gold, L. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6988.
(84) Kohn, K. W. *Development of Target-Oriented Anticancer Drugs*; Cheng, Y.-C.; Goz, B.; Minkoff, M.; Eds.; Raven: New York, 1983; p 181.
(85) Wemmer, D. E.; Benight, A. S. *Nucleic Acids Res.* **1985**, *13*, 8611.
(86) Erie, D. A.; Jones, R. A.; Olson, W. K.; Sinha, N. K.; Breslauer, K. J. *Biochemistry* **1989**, *28*, 268.
(87) Ashley, G. W.; Kushlan, D. M. *Biochemistry* **1991**, *30*, 2927.
(88) Bannwarth, W.; Dorn, A.; Iaiza, P.; Pannekouke, X. *Helv. Chim. Acta* **1994**, *77*, 182.
(89) Hirao, I.; Kawai, G.; Yoshizawa, S.; Nishimura, Y.; Ishido, Y.; Watanabe, K.; Miura, K. *Nucleic Acids Res.* **1994**, *22*, 576.
(90) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; d'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I., Jr.; Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1364.
(91) Durand, M.; Chevrie, K.; Chassignol, M.; Thuong, N. T.; Maurizot, J. C. *Nucleic Acids Res.* **1990**, *18*, 6353.
(92) Ma, M. Y.-X.; McCallum, K.; Climie, S. C.; Kuperman, R.; Lin, W. C.; Summer-Smith, M.; Barnett, R. W. *Nucleic Acids Res.* **1993**, *21*, 2585.
(93) Kool, E. T. *J. Am. Chem. Soc.* **1991**, *113*, 6265.
(94) Letsinger, R. L.; Wu, T. *J. Am. Chem. Soc.* **1995**, *117*, 7323.
(95) Wallace, T. J. *J. Org. Chem.* **1966**, *31*, 3071.
(96) Whitesides, G. M.; Houck, J.; Patterson, M. A. K. *J. Org. Chem.* **1983**, *48*, 112.
(97) Zuckermann, R.; Corey, D.; Schultz, P. *Nucleic Acids Res.* **1987**, *15*, 5305.
(98) Gupta, K. C.; Sharma, P.; Sathyanarayana, S.; Kumar, P. *Tetrahedron Lett.* **1990**, *31*, 2471.
(99) Johnson, R. E.; Adams, P.; Rupley, J. A. *Biochemistry* **1978**, *17*, 1479.
(100) Creighton, T. E. *BioEssays* **1988**, *8*, 57.
(101) Lipsett, M. N. *Cold Spring Harb. Symp. Quant. Biol.* **1966**, *31*, 449.
(102) Lipsett, M. N. *J. Biol. Chem.* **1967**, *242*, 4067.
(103) Lipsett, M. N.; Doctor, B. P. *J. Biol. Chem.* **1967** *242*, 4072.
(104) Ferentz, A. E.; Verdine, G. L. *J. Am. Chem. Soc.* **1991**, *113*, 4000.
(105) Ferentz, A. E.; Keating, T. A.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 9006.
(106) Erlanson, D. A.; Chen, L.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 12583.
(107) Wolfe, S. A.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 12585.
(108) Allerson, C. R.; Verdine, G. L. *Chem. Biol.* **1995**, *2*, 667.
(109) Wolfe, S. A.; Ferentz, A. E.; Grantcharova, V.; Churchill, M. E. A.; Verdine, G. L. *Chem. Biol.* **1995**, *2*, 213.
(110) Glick, G. D. *J. Org. Chem.* **1991**, *56*, 6746.
(111) Glick, G. D.; Osborne, S. E.; Knitt, D. S.; Marino, J. P., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 5447.
(112) Goodwin, J. T.; Glick, G. D. *Tetrahedron Lett.* **1994**, *35*, 1647.
(113) Goodwin, J. T.; Osborne, S. E.; Swanson, P. C.; Glick, G. D. *Tetrahedron Lett.* **1994**, *35*, 4527.
(114) Goodwin, J. T.; Osborne, S. E.; Scholle, E. J.; Glick, G. D. *J. Am. Chem. Soc.* **1996**, *118*, 5207.
(115) Wang, H.; Osborne, S. E.; Zuiderweg, E. R. P.; Glick, G. D. *J. Am. Chem. Soc.* **1994**, *116*, 5021.
(116) Wang, H.; Zuiderweg, E. R. P.; Glick, G. D. *J. Am. Chem. Soc.* **1995**, *117*, 2981.
(117) Osborne, S. E.; Völker, J.; Stevens, S. Y.; Breslauer, K. J.; Glick, G. D. *J. Am. Chem. Soc.* **1996**, *118*, 11993.
(118) Völker, J.; Osborne, S. E.; Glick, G. D.; Breslauer, K. J. *Biochemistry* **1997**, *36*, 756.
(119) Milton, J.; Connolly, B. A.; Nikforov, T. T.; Cosstick, R. *J. Chem. Soc., Chem. Commun.* **1993**, 779.
(120) Gao, H.; Yang, M.; Cook, A. F. *Nucleic Acids Res.* **1995**, *23*, 285.
(121) Gao, H.; Yang, M.; Patel, R.; Cook, A. F. *Nucleic Acids Res.* **1995**, *23*, 2025.
(122) Chaudhuri, N. C.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 10434.
(123) Sigurdsson, S. T.; Tuschl, T.; Eckstein, F. *RNA* **1995**, *1*, 575.
(124) Nelson, J. S.; Giver, L.; Ellington, A. D.; Letsinger, R. L. *Biochemistry* **1996**, *35*, 5339.
(125) Osborne, S. E.; Ellington, A. D. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2339.
(126) Jellinek, D.; Green, L. S.; Bell, C.; Janjic, N. *Biochemistry* **1994**, *33*, 10450.
(127) Li, W.-X.; Kaplan, A. V.; Grant, G. W.; Toole, J. J.; Leung, L. L. K. *Blood* **1994**, *83*, 677.
(128) Griffin, L. C.; Tidmarsh, G. F.; Bock, L. C.; Toole, J. J.; Leung, L. L. K. *Blood* **1993**, *81*, 3271.
(129) Hale, S. P.; Schimmel, P. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2755.
(130) Symensma, T. L.; Giver, L.; Zapp, M.; Takle, G. B.; Ellington, A. D. *J. Virology* **1996**, *70*, 179.
(131) Good, P. D.; Kirkos, A. J.; Li, S. X. L.; Bertrand, E.; Lee, N. S.; Giver, L.; Ellington, A. D.; Zaia, J. A.; Rossi, J. J.; Engelke, D. R. *Gene Therapy*, in press.
(132) Trubetskoy, V. S.; Torchilin, V. P.; Kennel, S.; Huang, L. *Biochim. Biophys. Acta* **1992**, *1131*, 311.
(133) Bunnell, B. A.; Askari, F. K.; Wilson, J. M. *Somatic Cell Mol. Genet.* **1992**, *18*, 559.
(134) Litzinger, D. C.; Huang, L. *Biochim. Biophys. Acta* **1992**, *1113*, 201.
(135) Lurquin, P. F. *Liposome Technology;* CRC: Boca Raton, FL, 1993; Vol. II, p 129.
(136) Felgner, P. L. *Human Gene Therapy* **1996**, *7*, 1791.
(137) Wang, S.; Lee, R. J.; Cauchon, G.; Gorenstein, D. G.; Low, P. S. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3318.
(138) Stein, C. A.; Tonkinson, J. L.; Zhang, L.-M.; Yakubov, L.; Gervasoni, J.; Taub, R.; Rotenberg, S. A. *Biochemistry* **1993**, *32*, 4855.
(139) Loke, S. L.; Stein, C. A.; Zhang, X. H.; Mori, K.; Nakanishi, M.; Subasinghe, C.; Cohen, J. S.; Neckers, L. M. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 3474.
(140) Geselowitz, D. A.; Neckers, L. M. *Antisense Res. Dev.* **1992**, *2*, 17.
(141) Marti, G.; Egan, W.; Noguchi, P.; Zon, G.; Matsukura, M.; Broder, S. *Antisense Res. Dev.* **1992**, *2*, 27.
(142) Iverson, P. L.; Zhu, S.; Meyer, A.; Zon, G. *Antisense Res. Dev.* **1992**, *2*, 211.
(143) Stein, C. A.; Cheng, Y.-C. *Science* **1993**, *261*, 1004.
(144) Chin, D. J.; Green, G. A.; Zon, G.; Szoka, F. C.; Straubinger, R. M. *New Biol.* **1990**, *2*, 1091.
(145) Bartel, D. P.; Zapp, M. L.; Green, M. R.; Szostak, J. W. *Cell* **1991**, *67*, 529.
(146) Zapp, M. L.; Stern, S.; Green, M. R. *Cell* **1993**, *74*, 969.
(147) Baskerville, S.; Zapp, M.; Ellington, A. D. *J. Virol.* **1995**, *69*, 7559.
(148) Rieger, A.; Nassal, M. *Nucleic Acids Res.* **1995**, *23*, 3909.
(149) Famulok, M.; Huttenhofer, A. *Biochemistry* **1996**, *35*, 4265.
(150) Wallis, M. G.; von Ahsen, U.; Schroeder, R.; Famulok, M. *Chem. Biol.* **1995**, *2*, 543.
(151) Swanson, H. I.; Chan, W. K.; Bradfield, C. A. *J. Biol. Chem.* **1995**, *270*, 26292.
(152) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1996**, *3*, 369.

(153) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160.
(154) Singh, R.; Valcarcel, J.; Green, M. R. *Science* **1995**, *268*, 1173.
(155) Tacke, R.; Manley, J. L. *EMBO J.* **1995**, *14*, 3540.
(156) Ringquist, S.; Jones, T.; Snyder, E. E.; Gibson, T.; Boni, I.; Gold,. L. *Biochemistry* **1995**, *34*, 3640.
(157) Kubik, M. F.; Stephens, A. W.; Schneider, D.; Marlar, R. A.; Tasset, D. *Nucleic Acids Res.* **1994**, *22*, 2619.
(158) Wang, Y.; Rando, R. R. *Chem. Biol.* **1995**, *2*, 281.
(159) Lato, S. M.; Boles, A. R.; Ellington, A. D. *Chem. Biol.* **1995**, *2*, 291.
(160) Lato, S. M.; Ellington, A. D. *Mol. Div.* **1996**, *2*, 103.
(161) Ellington, A. D.; Szostak, J. W. *Nature* **1992**, *355*, 850.
(162) Lauhon, C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 1246.
(163) Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550.
(164) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425.
(165) Ciesiolka, J.; Gorski, J.; Yarus, M. *RNA* **1995**, *1*, 538.
(166) Ciesiolka, J.; Yarus, M. *RNA* **1996**, *2*, 785.
(167) Li, Y.; Geyer, R.; Sen, D. *Biochemistry* **1996**, *35*, 6911.
(168) Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1994**, *33*, 973.
(169) Burgstaller, P.; Famulok, M. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1084.
(170) Fan, P.; Suri, A. K.; Fiala, R.; Live, D.; Patel, D. J. *J. Mol. Biol.* **1996**, *258*, 480.
(171) Burgstaller, P.; Kochoyan, M.; Famulok, M. *Nucleic Acids Res.* **1995**, *23*, 4769.
(172) Yang, Y.; Kochoyan, M.; Burgstaller, P.; Westhof, E.; Famulok, M. *Science* **1996**, *272*, 1343.
(173) Jiang, F.; Kumar, R. A.; Jones, R. A.; Patel, D. J. *Nature* **1996**, *382*, 183.
(174) Ellington, A. D. *Drug Develop. Res.* **1994**, *33*, 102.
(175) Lown, J. W. *Chemtracts Org. Chem.* **1993**, *6*, 205.
(176) Nielsen, P. E. *Bioconjugate Chem.* **1991**, *2*, 1.
(177) Leclerc, F.; Cedergren, R.; Ellington, A. D. *Nature Struct. Biol.* **1994**, *1*, 293.
(178) Wiegand, T. W.; Williams, P. B.; Dreskin, S. C.; Jouvin, M.-H.; Kinet, J.-P.; Tasset, D. *J. Immunol.* **1996**, *157*, 221.

CR960009C