Part 1
History and Theoretical Background

The Aptamer Handbook. Edited by S. Klussmann

ISBN 3-527-31059-2

1
In Vitro Selection of Functional Oligonucleotides and the Origins of Biochemical Activity

James M. Carothers and Jack W. Szostak

1.1
Introduction

In vitro selection is an experimental method for searching oligonucleotide sequence spaces for synthetic structures and activities. Oligonucleotide sequence spaces are very large – they contain the ensemble of all possible sequences of a given length separated by point mutations (Maynard Smith, 1970). For example, the sequence space of an RNA the length of a small tRNA (74 nucleotides) encompasses 10^{43} different molecules. The largest libraries typically synthesized in the laboratory, approximately 10^{16} different sequences, represent only a minute fraction of the total number of possible sequences for any nucleic acid molecule of even modest size (Wilson and Szostak, 1999). How can such necessarily sparse samplings of sequence space produce so many different aptamers, ribozymes, and deoxyribozymes? In this chapter, we focus on the technology of *in vitro* selection and what its application teaches us about the quantity and quality of functional structures in nucleic acid sequence spaces.

We begin with a basic introduction to *in vitro* selection in Section 1.2, presenting a brief history of the technology growing out of the discovery of naturally occurring ribozymes. Section 1.3 deals with the synthetic aptamers, ribozymes, and deoxyribozymes generated in the last 15 years. Rather than enumerating all the successes (see Wilson and Szostak, 1999; other chapters this volume), we condense the results into a set of general lessons about the capabilities and potential limitations of functional nucleic acids. We pay particular attention to the distribution of activity in sequence space. In the next section (Section 1.4) we consider the origins of natural biochemical activity. We elaborate on some of the basic questions about the composition of simple life forms and describe how synthetic approaches make these problems accessible to experimentation. In the last section (Section 1.5) we highlight recent technological developments and research ambitions.

The Aptamer Handbook. Edited by S. Klussmann

ISBN 3-527-31059-2

1.2 A Brief History of *In Vitro* Selection

Suggestions came as early as the 1960s that in addition to their templating functions, nucleic acids could fold into complex three-dimensional shapes and perform biochemical activities (Benner et al., 1999). By the 1980s, work from the Cech, Altman, and Pace groups demonstrated that RNA can catalyze at least two different kinds of chemical reactions (reviewed in Cech, 2002). Cech (Kruger et al., 1982) identified the *Tetrahymena thermophila* group I intron as a naturally occurring RNA that catalyzes a transesterification reaction. Likewise, the RNA portion of RNase P was shown to be both necessary and sufficient for the catalysis of the hydrolysis of the correct phosphodiester bond of its pre-tRNA substrates (Gardiner and Pace, 1980; Guerrier-Takada et al., 1983).

The discovery of ribozymes renewed speculation about the functional capabilities of RNA (Cech, 1986; Orgel, 1986). Gilbert coined the phrase "RNA world" (Gilbert, 1986) to refer to an imagined stage in the early evolution of life where RNA carried out most of the biochemical functions in the cell. In this model, ribozymes with RNA polymerase activity replicated all of the functional RNA structures in the cell. The RNA world hypothesis offered a conceptual solution to the problem of how the original functional biochemical structures could have been copied, but initiated a host of other questions. Chief among them, how did the self-replicating RNAs arise? At the time, estimates of the probability of finding an RNA in a pool of random molecules with *any* binding or catalytic activity, much less polymerase function, ranged from the relatively likely 1 in 10^5 to the practically impossible 1 in 10^{50} (Lorsch and Szostak, 1996).

Another important unanswered question was whether the functional capabilities of RNA are diverse and robust enough for RNA to have performed all of the functions necessary in a simple cell. Rational design showed that natural ribozymes can be remodeled to perform their chemistries in different contexts. Been and Cech (1988) achieved primer extension by a modified *Tetrahymena* group I intron; Doudna and Szostak (1989) successfully altered the group I intron to splice together strands of RNA on an exogenous template; Pace and co-workers used phylogenetic analysis to guide the creation of a minimized RNase P catalytic subunit (Waugh et al., 1989). Although those early engineering efforts were successful, it became clear that strategies based on laboratory evolution would be better suited to address questions about the range of functions that RNA can perform and the abundance of active sequences in collections of random sequences (Szostak, 1988; Joyce, 1989).

In the classic test tube evolution experiments conducted by Sol Spiegelman, superior variants were isolated from populations of sequences undergoing mutation at the rate inherent to the Qβ RNA polymerase (Saffhill et al., 1970; Wilson and Szostak, 1999). Modern *in vitro* selection experiments take advantage of the ability to direct the chemical synthesis of large combinatorial libraries. Oligonucleotide pools containing as many as 10^{16} different sequences can be created with nearly any design. The basic scheme for *in vitro* selection is outlined in Fig. 1.1.

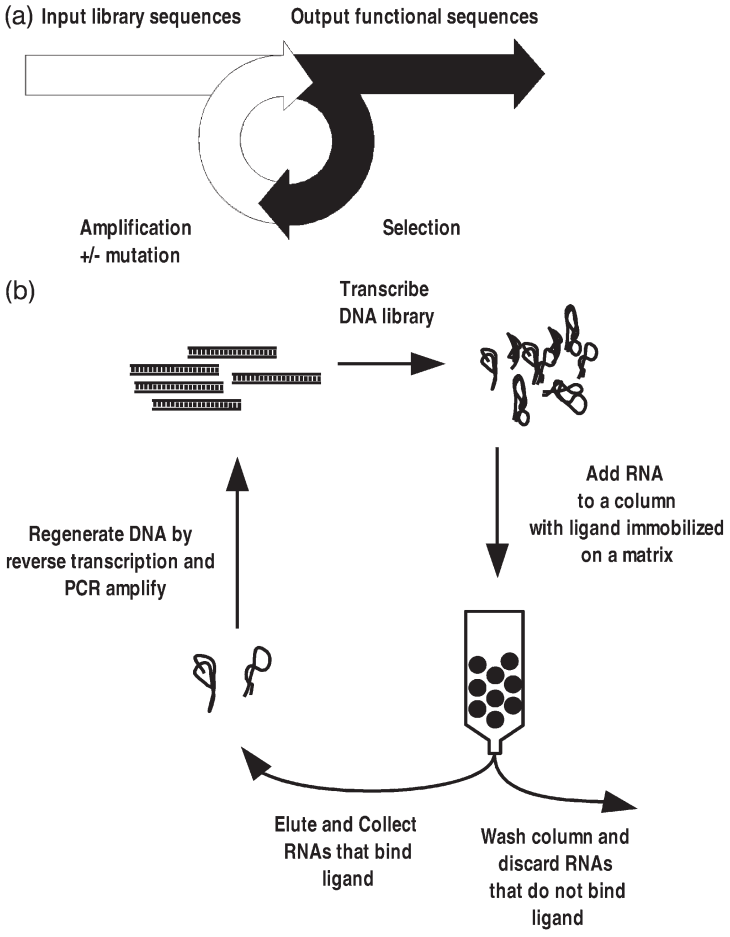


Fig. 1.1 Schemes for *in vitro* selection. (a) General scheme for *in vitro* selection. *In vitro* selection is an iterative method for searching libraries of partially random or completely random sequences for molecules that can perform a given function. Sequences from an input library are subjected to a selective step in which active molecules are separated from inactive molecules. Sequences that survive the selective step are amplified by polymerase chain reaction (PCR). The cycle is repeated until functional molecules dominate the population, at which point they are cloned and sequenced. (b) Scheme for an RNA aptamer selection. Transcribing a library of DNA molecules produces RNA. In this example, RNAs that do not bind to the ligand immobilized on the column matrix are washed away. RNAs that bind the column and specifically elute are collected. Reverse transcription and PCR generate a pool of DNA whose corresponding RNA sequences are enriched in ligand-binding activity.

By 1990, *in vitro* selection was used to isolate rare RNAs from random sequence pools that could bind arbitrarily chosen targets (Ellington and Szostak, 1990; Tuerk and Gold, 1990). Ellington and Szostak termed the resulting motifs "aptamers," while Tuerk and Gold dubbed the experimental process itself "systematic

evolution of ligands by exponential amplication" (SELEX). Other applications of the technology soon followed. The same kind of phylogeny-directed analysis conducted on RNase P was demonstrated with the *Tetrahymena* ribozyme, where functional sequence variants were generated and identified using *in vitro* selection (Green et al., 1990). *Tetrahymena* ribozymes that could cleave a DNA substrate were readily isolated (Robertson and Joyce, 1990). Selections for single-stranded DNA molecules that bind to small organic dyes (Ellington and Szostak, 1992) and thrombin (Bock et al., 1992) proved that oligonucleotides other than RNA can also form complex shapes that exhibit non-templating functions.

RNA aptamers to adenosine triphosphate (ATP) were isolated from random sequence pools by immobilizing the ligand on a solid support followed by specific elution with ATP in solution (Sassanfar and Szostak, 1993). Yeast $tRNA^{Phe}$ had previously been shown to be a Pb^{2+}-dependent self-cleaving RNA (Brown et al., 1983). An effort to identify molecules with improved lead-dependent self-cleaving activity from a pool of mutagenized $tRNA^{Phe}$ sequences resulted in diverse novel structures (Pan and Uhlenbeck, 1992).

The first entirely synthetic ribozymes were molecules with RNA–RNA ligation activity (Bartel and Szostak, 1993). Several different classes of molecules were obtained from a library of 10^{15} different sequences each with 220 random positions. In a manner reminiscent of the Qβ polymerase experiments, error-prone polymerase chain reaction (PCR) was used to generate mutations in later rounds of the ligase selection in order to search areas in sequence space that surrounded the initial isolates. Another technique, DNA shuffling, was created to recombine sequence variations that would otherwise not be likely to be found in the same molecule, further increasing the amount of sequence space that can be searched (Stemmer, 1994; Zhao and Arnold, 1997). Systems of continuous evolution (Breaker et al., 1994) offer a way to perform dozens of rounds of selection in a relatively short time period. Population behaviors such as cooperativity and parasitism could be observed once continuous evolution was applied to ribozyme selections (Wright and Joyce, 1997; reviewed by Robertson and Ellington, 1997).

1.3 Lessons from the Aptamers, Ribozymes, Deoxyribozymes Generated by *In Vitro* Selection

What have we learned from *in vitro* selection about the functional capabilities of oligonucleotides? Fast forward to the present and one finds that synthetic RNAs have been identified that can perform many different kinds of reactions including phosphodiester bond formation and cleavage, carbon–carbon bond formation, alkylation, a Diels-Alder condensation and numerous acyl-transfer reactions including amide bond formation, among others (Joyce, 2002). RNA aptamers have been selected to hundreds of small molecule and protein targets including, for example, ATP, GTP, B12, malachite green and caffeine, HIV-1 Rev peptide, MS2 coat protein, and thrombin (reviewed in Wilson and Szostak, 1999 and Hermann

and Patel, 2000). Biochemical and structural characterization (for a review of aptamer structures see Hermann and Patel, 2000 shows that synthetic RNA structures can be stable, intricate, and exhibit very specific binding or catalytic activity. Taken together, these results demonstrate that many different functional RNA structures exist and that these can be accessed through relatively sparse searches of sequence space.

In light of all the successes one might ask if there is anything that RNA *cannot* do. Given the challenge of keeping free radicals localized so they do not cause harm, chemical mechanisms that involve radicals may be difficult for RNA (Joyce, 2002). However, as with naturally occurring proteins, oligonucleotides can use cofactors to introduce structural or chemical characteristics they do not possess (for example, Roth and Breaker, 1998; Tsukiji et al., 2004). With the right cofactor, perhaps B12 or a quinone, a ribozyme or deoxyribozyme may be able to catalyze a radical-dependent reaction such as ribonucleotide reduction (Stubbe, 2000).

Because the phosphodiester backbone of oligonucleotides is polyanionic, counterions are required for proper folding (Lilley, 2003). The metal requirements, particularly for divalent cations, can be very specific. For instance, the class I ligase ribozyme, as it was originally selected, requires Mg^{2+} (Glasner et al., 2002), while other *in vitro* selected oligonucleotides need Pb^{2+} (Pan and Uhlenbeck, 1992) or Ca^{2+} (Okumoto et al., 2003). On the other hand, structures with altered specificity for divalent cations can be selected (Lehman and Joyce, 1993; Riley and Lehman, 2003). And there are examples where divalent cations are dispensable even though the RNA was selected in conditions with a high concentration of the metal (Dieckmann et al., 1996).

The widespread success in the selection of functional RNAs raised the issue of whether related nucleic acids could also give rise to functional sequences. Single-stranded DNA aptamers have been selected for many different targets (Wilson and Szostak, 1999). Likewise, deoxyribozymes can catalyze a variety of phosphodiester bond-oriented reactions. In principle, DNA should be able to catalyze as many different kinds of chemistries as RNA, especially with the help of metals or other cofactors (Li and Sen, 1997; Sidorov et al., 2004; Silverman, 2004). With few exceptions, (for example, Lauhon and Szostak, 1995) DNAs made of the same sequence as a functional RNA motif are not active (Huizenga and Szostak, 1995). And yet, the lack of a 2′OH does not seem to diminish the ability of DNA to form aptamers or enzymes.

Clearly oligonucleotides such as RNA and DNA can perform many simple biochemical tasks. And, interestingly, they do not have to be made using only, or even all of, the four canonical nucleotides (A, G, C, U/T). Selections with random libraries comprising reduced sets of nucleotides reveal that molecules with even less chemical diversity can form functional structures. Rogers and Joyce (1999) found that ligase ribozymes could be built with only three bases, A, G, and U. Even a ribozyme with only two types of nucleotides, 2,6-diaminopurine and uracil can catalyze a 5′–3′ RNA–RNA ligation 36 000 times faster than background (Reader and Joyce, 2002). Note, however, that the rate enhancement observed

for the ribozyme with two kinds of nucleotides was very slow compared with ligases made of three or four base types (rate enhancements of 10^5 and $>10^6$, respectively) (Reader and Joyce, 2002; Bartel and Szostak, 1993).

Selections from random libraries with modified nucleotides demonstrate that the "heteropolymer space" surrounding RNA and DNA also contains active molecules. Vaish et al. (2003) made libraries which incorporated a cationic group, a functionality not usually found in RNA. ATP-binding aptamers were generated whose activity depended on the presence of the modification. Yet, a side by side selection for ATP aptamers from pools with and without a different cationic modification yielded the same weakly binding motif (Battersby et al., 1999). This suggests that at the level of binding stringency required to survive the Battersby selection, sequences possessing the cationic moiety may have been at a disadvantage relative to those made up of the ordinary RNA bases only. Replication bias may have resulted in such a low level of the analog base that the composition of the pool was effectively limited to the three ordinary nucleotides. Synthetic ribozymes whose sugars have 2′-*O*-Me, 2′-F, or 2′-NH_2 modifications are frequently used for therapeutic or technological applications because of their reduced degradation rates (Zinnen et al., 2002). In a similar vein, many naturally occurring RNAs have modified nucleotides such as pseudouridine or 2′-*O*-Me sugars that improve their folding and stability (Eliceiri, 1999).

Selections typically lead to the emergence of the simplest and therefore most common motif capable of surviving the pressure applied (Wilson and Szostak, 1999). The effect is that the same molecule can be discovered repeatedly if the challenges and conditions are the same even though other, more complicated, solutions to the problem may exist. For example, the same "Sassanfar" ATP RNA aptamer (Sassanfar and Szostak, 1993) was isolated in different labs over the course of several years in response to selection against adenosine-containing ligands (Burgstaller and Famulok, 1994; Burke and Gold, 1997). When the selection was repeated with the additional requirement that the aptamers discriminate between binding ATP and AMP, a new motif was recovered (Sazani et al., 2004). Evidently the new aptamer requires more sequence information to specify its structure than the Sassanfar aptamer. A given pool of random molecules should contain more sequences corresponding to the simpler Sassanfar motif than the new, more complex aptamer. Accordingly, without selection pressure against it, the Sassanfar motif dominates the pool of active molecules. The hammerhead ribozyme may be a natural example of the same phenomenon. *In vitro* selection for cleavage activity in near-physiological conditions produced the same hammerhead motif found in diverse natural phyla (Salehi-Ashtiani and Szostak, 2001).

Pragmatism motivates interest in the kinds of library designs most likely to yield functional oligonucleotides (Bartel and Szostak, 1993; Sabeti et al., 1997; Davis and Szostak, 2002). The hypothesis that longer sequence pools are better sources of activity than shorter pools relies on two rationales: (1) Larger oligonucleotides have an inherently greater capacity to fold into functional conformations than shorter molecules. (2) Complex molecules can be split into smaller modules where the likelihood of finding the motifs together in the same sequence in-

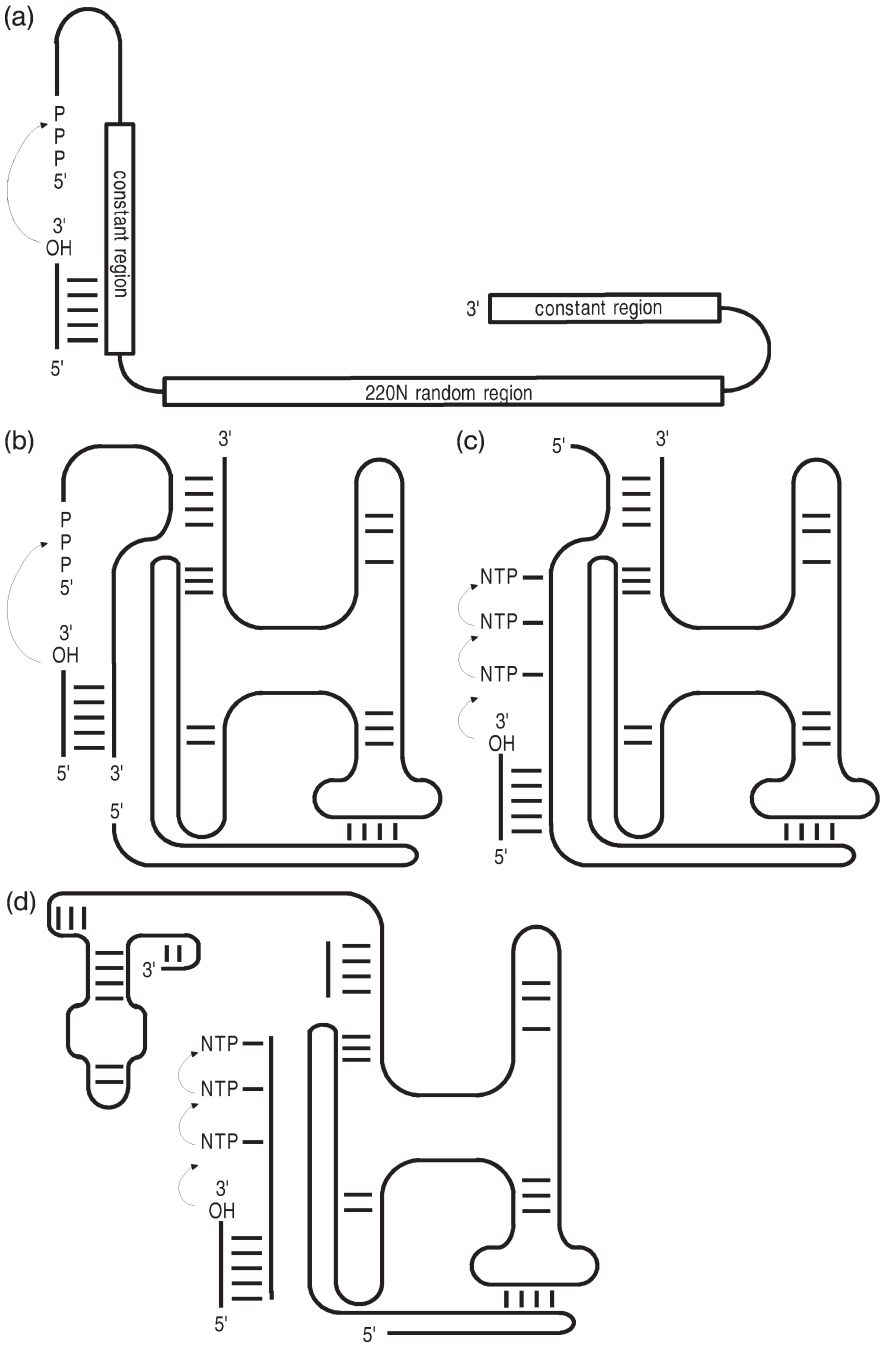
creases with the length of the pool. Calculations show that the potential advantages of larger pools depend highly on the number of modules into which the functional structure can be broken (Knight and Yarus, 2003). When sequences in a random library are longer than the functional structure, the functional motif can be present in many different positions, or registers, within a pool molecule. If a functional structure (N nucleotides long) is shorter than the pool molecules (L nucleotides long), the likelihood of finding the functional structure increases by a factor of $(L-N)$. When the functional structure can be assembled from stretches of sequence interspersed with unconstrained regions of variable lengths, further increases in the abundance of active molecules are attained.

The theoretical benefits of long random oligonucleotide pools are balanced by the realities that long pools tend to aggregate (Bartel and Szostak, 1993) and that otherwise functional structures may have a propensity to misfold when extraneous stretches of sequence are present (Sabeti et al., 1997). The negative aspects of using pools longer than 70 or 80 random nucleotides imply that they should only be used when the goal is to obtain binding or catalytic activities expected to require very large, very complex structures.

Ultimately, the answer to the question of how to best design sequence libraries depends on the goal, the underlying fitness landscapes and the distribution of active structures in sequence space. In some cases there may be simple evolutionary pathways that connect a structure with one activity to a structure with a different activity (Hanczyc and Dorit, 2000; Lehman et al., 2000; Reidys et al., 2001. These structures may lie surprisingly close to one another in sequence space. For example, Schultes and Bartel (2000) found two ribozymes with entirely different secondary structures and very different functions that nonetheless share common sequences that form a neutral mutational "pathway of connectivity" between them (see also Held et al., 2003).

In some cases, new functionality can be evolved from an existing structure without the shape of the molecule undergoing substantial rearrangement. The class I ligase from the Bartel and Szostak selection (1993) was converted first into a multiple turnover enzyme (Ekland et al., 1995) and then into a ribozyme with template-directed primer-extension activity (Ekland and Bartel, 1996). Eventually, a new domain was added to the 3′ end of the core structure that improves the primer-extension activity to the point that 14 bases can be polymerized in a sequence-dependent manner (Johnston et al., 2001; see Fig. 1.2). The mechanistic similarities between RNA–RNA ligation and RNA polymerization likely facilitated the conversion. The ligase and the polymerase both recognize a double-stranded RNA duplex and position the α-phosphate of a downstream nucleotide for attack by a 3′ OH.

Even with high-resolution structural data in hand, it can be difficult to predict whether a given functional molecule is a good starting point for the evolution of a new activity. Side by side selections revealed that a library containing an ATP aptamer flanked by random sequence was not a better source of ATP-dependent nucleotide kinase ribozymes than a completely random pool (Urbach, 2000). Because kinases need to position the γ-phosphate of ATP in order to activate it



◀ **Fig. 1.2** Stepwise progression from the class I ligase ribozyme to an RNA-dependent RNA polymerase with primer extension activity. (a) The class I ligase was isolated from a library of 10^{15} different 220 nucleotide-long sequences. It catalyzes the formation of a phosphodiester bond between the 3′ hydroxyl of the substrate and its own 5′ triphosphate. Base-pairing aligns the substrate with the ribozyme (Bartel and Szostak, 1993). (b) Selection from a mutagenized library based on the class I ligase-identified functional sequence variants. The sequence variation data were used to generate a secondary structure model and guide the design of a molecule with multiple-turnover, intermolecular ligation activity (Ekland and Bartel, 1995; Ekland et al., 1995). (c) Further engineering converted the ligase into a ribozyme that extends a primer along an internal template by incorporating up to six nucleotides (Ekland and Bartel, 1996). (d) An extra random sequence domain was appended to the catalytic core of the ribozyme. Functional sequences were selected that use an external template to extend a primer up to 14 nucleotides (Johnston et al., 2001).

for nucleophilic attack, the expectation was that library sequences with the embedded ATP aptamer would have a selective advantage relative to completely random molecules. The nuclear magnetic resonance (NMR) structure of the aptamer shows that the ATP is stably bound and that there is good solvent accessibility for the γ-phosphate (Jiang et al., 1996; Dieckman et al., 1996). Perhaps the γ-phosphate enjoyed too much conformational freedom to provide an advantage to the aptamer-containing library. Or it may be that structures compatible with the geometry of that particular ATP aptamer are, by chance, very rare in sequence space.

Several aptamer structures have been shown to be readily evolvable in terms of specificity. Three mutations within the binding loop of an *in vitro* selected L-citrulline aptamer were enough to change its specificity to L-arginine because the pattern of hydrogen bond donors and acceptors could be flipped in a simple way (Yang et al., 1996). Modifying a single base from C to U alters the specificity of the binding domains of naturally occurring riboswitches from guanine to adenine (Mandal and Breaker, 2004). In other instances, the evolution of new binding specificity has resulted in an entirely different fold. Structures with completely new secondary (and presumably) tertiary configurations were obtained upon evolving the specificity of three aptamers from flavin-adenine diphosphate (FAD) to guanosine monophosphate (GMP), even though the closest overlapping solutions differ by only three sequence mutations (Held et al., 2003). Similarly, selection for GTP binders from a pool of mutagenized ATP aptamers produced molecules with highly diverged sequences and secondary structures (Huang and Szostak, 2003).

In a broad sense, it is very difficult to know whether a new function can be evolved from a pre-existing structure. However, there are methods of designing libraries that increase the chances of producing active structures. Oligonucleotides have a tendency to collapse into folded states because of the ease with which Watson–Crick and wobble-type pairings form (Wilson and Szostak, 1999). Not surprisingly, functional oligonucleotides usually contain one or more simple secondary structural motifs such as stems, stable tetraloops, pseudoknots,

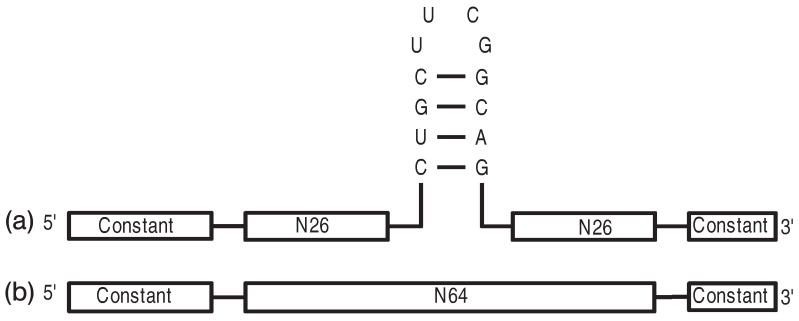


Fig. 1.3 Partially engineered RNA sequence libraries. In one successful example, an RNA library consisting of (a) 2.5×10^{14} random sequences containing a stable internal stem–tetraloop ("partially engineered") and (b) 2.5×10^{14} completely random sequences ("undesigned") was used to select high-affinity aptamers for GTP (Davis and Szostak, 2002). The partially engineered sequences proved to be a much better source of high-affinity aptamers than the undesigned fully random pool (also see Fig. 1.4).

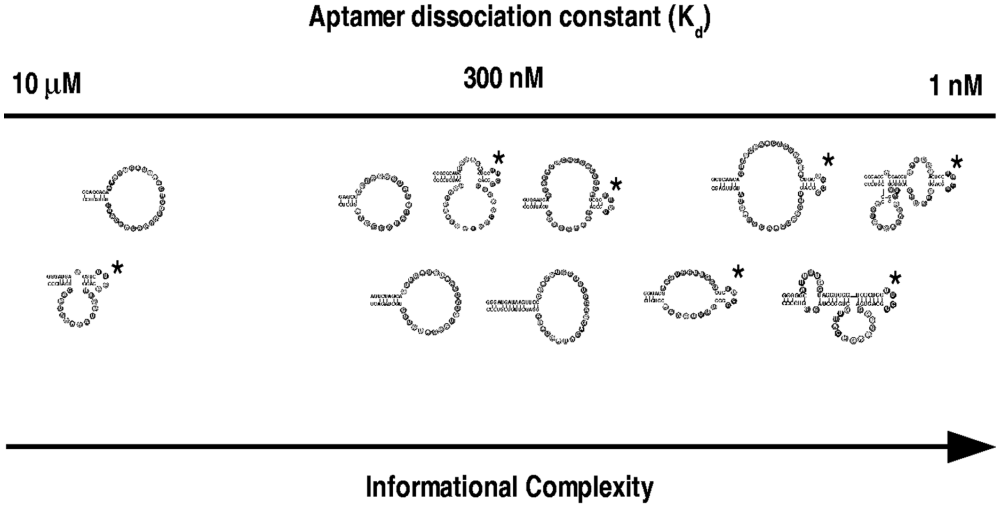


Fig. 1.4 Informational complexity and functional activity. More information is required to specify aptamers that bind a ligand more tightly (Carothers et al., 2004). In this figure, the secondary structure of each of a series of GTP-binding aptamers is placed according to its dissociation constant. Moving to the right (tighter binding), the aptamers have more intricate secondary structures and are more informationally complex. An asterisk indicates structures that originated from the partially engineered sequence library. Notably seven of the 11 aptamers came from the partially engineered portion of the library, including six of the seven aptamers with dissociation constants better than 300 nmol/L.

and hairpins. It may be better to build libraries with random sequence flanking these kinds of elements, than to use completely random molecules or motifs preselected for a particular activity. In one successful example, a library of 2.5×10^{14} molecules was constructed with an internal stem and stable tetraloop flanked by

26 completely random bases on each side (see Fig. 1.3; Davis and Szostak, 2002). The partially designed library was mixed with 2.5×10^{14} completely random sequences of the same length and subjected to selection for high-affinity binding to GTP. Eleven different classes of aptamers with dissociation constants (K_d values) ranging from 8 μmol/L to 9 nmol/L were isolated and characterized (Davis and Szostak, 2002; Carothers et al., 2004). Seven of the 11 aptamers came from the partially designed pool, including, remarkably, six of the seven with affinities better than 300 nmol/L (see Fig. 1.4).

To begin to understand how the abundance of functional structures in sequence space varies with the difficulty of the biochemical problem, we determined experimentally the amount of information (Schneider et al., 1986; Adami and Cerf, 2000; Adami, 2004) required to specify each of the 11 GTP aptamer structures in the conditions used in the Davis and Szostak selection (Carothers et al., 2004). *In vitro* selection was used to identify functional sequence variants for each of the aptamers. Roughly speaking, the information required to specify a structure increases as the number of functional sequence variants decreases (in other words, a structure that is rarer in sequence space requires more information to specify than a simpler structure that is more abundant). Based on the way that informational complexity varied with activity, each tenfold increase in binding affinity was shown to require an RNA structure that is about 1000 times less frequent in a pool of random sequences. Because the informational cost of the internal stem–tetraloop was "pre-paid," the biased design probably increased the likelihood of finding high-activity aptamers by 250- to 1000-fold.

Despite the presence of the engineered stem–tetraloop, several of the GTP aptamers are so informationally complex that they are extremely unlikely to be present in a pool with the length and composition used in the selection (Davis and Szostak, 2002; Carothers et al., 2004). One way to explain their presence, previously invoked to justify the discovery of the class I ligase (Ekland and Bartel, 1995), is that there are a very large number of different complex functional structures. Therefore, although the probability of finding any one particular complex structure may be very low, the number of different possibilities is large enough that a few informationally complex structures can be present in a diverse, well-designed pool.

In the case of the GTP aptamers, about ten more bits of information content, enough to specify five more conserved positions in an RNA molecule, are needed for each tenfold improvement in activity. Interestingly, the same quantitative relationship between activity and information was observed for a pair of ligase ribozymes (Carothers et al., 2004). Perhaps over the range of activities studied, the same types of structural changes can improve binding constants or rate constants. In both cases, increasing the thermodynamic stability of the active structures by forming more specific contacts within the RNA itself (Bergman et al., 2000) could lead to improvements in activity. If so, the informational cost of building structures with better activity could be comparable, regardless of the function. Analysis of the GTP aptamer binding specificities supports this hypothesis. Aptamers that exhibit tighter binding to GTP do not make better or more specific contacts with

the ligand (Oestreich, Carothers and Szostak, unpublished observations). Instead, tighter-binding aptamers have more intricate and stable secondary structures and greater propensities to fold into active conformations than weaker-binding molecules. These results demonstrate that, contrary to previous suggestions (Eaton et al., 1995), selection for high-affinity binding does not automatically produce aptamers with high specificity for the ligand.

1.4 Synthetic Approaches to Understanding the Natural Origins of Function

Biochemistry began on earth more than 3.5 billion years ago when chemical systems gave rise to simple biological systems that could grow, divide, and undergo Darwinian evolution (Mojzsis et al., 1999; Joyce, 2002). Understanding the origins of life is difficult, if for no other reason than because the distinction between living and non-living matter can be blurry (Szostak et al., 2001; Ruiz-Mirazo et al., 2004; McKay, 2004). At their simplest, living systems need an active metabolism to process inputs of mass, energy, and entropy (or information). And, to enable Darwinian evolution, they must inherit differences in survivability, growth, and reproduction. We imagine that a system needs to be relatively sophisticated in order to be considered alive, but much much simpler than any extant form of life.

Assuming the existence of the right precursors, was life likely to happen? The historical record does not provide many details about how early life forms worked; it only hints at what their compositions might have been (Simoneit, 2002). Deconstructing and minimizing existing organisms such as *Mycoplasma* (Smith et al., 2003; Check, 2002) will only tell us about cells containing DNA, RNA, and protein that are many times more complicated than the first were likely to have been.

An alternative approach is to "build up" simple synthetic structures and systems (Szostak et al., 2001, Luisi, 2002; Monnard, 2003; Hanczyc and Szostak, 2004) that are not burdened by the "non-functional" selective pressures in natural organisms (for example Zarrinpar et al., 2003). Synthetic methods can render otherwise intractable questions about the origins of biochemical function accessible to experimentation (Szostak et al., 2001; Luisi, 2002; Monnard and Deamer, 2002; Joyce, 2002; Benner, 2003). For example, (1) How easily can functional structures arise? (2) How many different biochemical structures does a simple organism need? (3) How and when do complexity and systems behavior emerge? (4) How does the nature of the chemical constituents affect evolutionary trajectories?

Life is a dissipative, open thermodynamic process. To some extent, the physical chemistry of life makes it similar to other systems, such as BZ chemical oscillators (reviewed in Schneider, 1985), or even fire (McKay, 2004). In all three, inputs of matter and/or energy enable the system to persist in a dynamic state far from equilibrium. But, among open thermodynamic systems, only life has the capacity for Darwinian evolution.

Among the early, but not necessarily earliest, ancestors of life on earth were discrete cell-like organisms with simple physical boundaries (Szostak et al., 2001;

Luisi, 2002; Ruiz-Mirazo et al., 2004). Physical boundaries such as membranes help maintain the concentration of other components and partially insulate the system from external fluctuations and parasites (Martin and Russel, 2003; Chen et al., 2004). There is much disagreement about whether metabolism, replicators, or some "emergent" network arose first (Kauffman, 1986; Martin and Russel, 2003; Chen et al., 2004; reviewed in Anet, 2004). For the purposes of this discussion, we are agnostic on that point and maintain that all of the models implicitly invoke complicated sets of reactions that eventually would have been catalyzed by heteropolymers.

Many chemical reactions are catalyzed efficiently without heteropolymers, or even macromolecules. For example, small non-polymeric biomimetic catalysts have been generated that selectively form asymmetric carbon–carbon bonds (reviewed in Cordova, 2004). However, to achieve *selective* reactivity in the presence of a complex mixture of potential reactants macromolecular catalysts are usually necessary (Breslow, 2001). Heteropolymeric macromolecules can provide extended complementarity not readily available in simpler catalysts, and the potential to generate new functionality from heteropolymers is certainly more open-ended than for other kinds of macromolecules (van Hest and Tirrell, 2001).

What were the earliest heteropolymers? What were the substrates for their synthesis? We focus attention not on determining the identity of those heteropolymers, but rather on the qualities that those molecules would have possessed (Fig. 1.5). To be biologically significant, a potential primordial heteropolymer

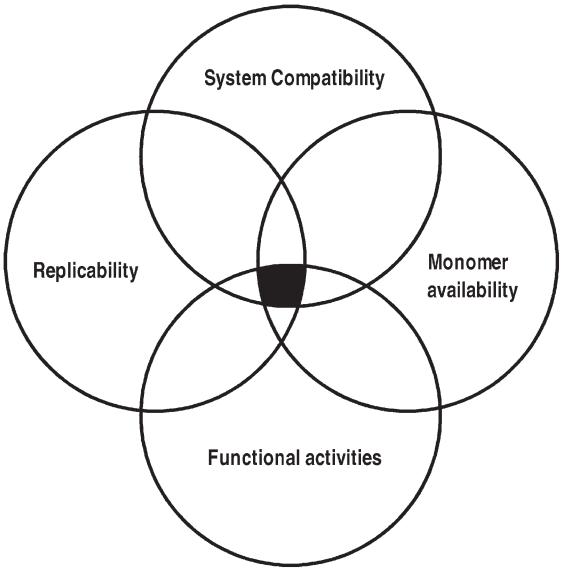


Fig. 1.5 Characteristics of the first biopolymers. To evaluate whether a given type of heteropolymer could have been the original biopolymer, we must consider at least four different properties. (Counterclockwise): There must be a simple mechanism for the heteropolymer to template its own synthesis (replicability). Functional structures that can bind to ligands or catalyze chemical reactions must be relatively abundant in the sequence space of the heteropolymer (functional activities). To synthesize the heteropolymer itself, there should be a simple, prebiotic route for generating its subunits (monomer availability). Finally, the conditions required for the functional heteropolymers to be stable and active should be compatible with other components likely to have been part of the system, such as membranes.

must have the following properties: (1) there must be a simple mechanism for templating the synthesis of additional heteropolymers from pre-existing ones, (2) there should be a chemically plausible supply of substrates necessary to synthesize the heteropolymer, (3) functional molecules must be accessible through relatively small searches of the sequence space of the heteropolymer, and (4) the heteropolymer and membrane components must be compatible with one another.

The first biopolymers are generally thought to more closely resemble RNA than protein because the base-pairing potential of oligonucleotides enables them to more readily serve as templates for their own synthesis (Joyce, 2002). With highly activated substrates, the templated synthesis of simple RNA sequences can be achieved even without enzymes (Joyce et al., 1984, 1987). Although occasionally invoked in thought experiments (for example, Godfrey-Smith, 2002), no simple physical mechanism for templating the open-ended synthesis of polypeptides from another polypeptide has been proposed or achieved (Cane and Walsh, 1999; Halpin et al., 2004).

A number of different oligonucleotides show base-pairing ability between complementary strands according to simple physical rules (reviewed in Eschenmoser 1999, 2004). Discerning a prebiotic, or even simple biogenic, route for the synthesis of the various monomer precursors can be complicated. Noteworthy in this regard, plausible prebiotic conditions leading to the synthesis and stabilization of pentose sugars, including ribose, were recently demonstrated (Ricardo et al., 2004).

RNA has the characteristics of a good model for the kind of heteropolymer the first life forms might have used even if it was not actually *the* original biopolymer (Benner et al., 1999). The capacity of RNA for highly developed function is exemplified by its presence at the heart of modern ribosomes (reviewed in Steitz and Moore, 2003) and spliceosomes (reviewed in Doudna and Cech, 2002). From the discussion in Section 1.3 (also see Joyce, 2002), we know that comparatively simple RNAs obtained from samplings of only 10^{12}–10^{15} different sequences, can exhibit many of the kinds of diverse functionalities a simple organism would have needed for its metabolism. *In vitro* selected ribozymes as small as 71 nucleotides can catalyze the synthesis of ribonucleotide monomers from activated ribose (pRpp) and pyrimidine (Unrau and Bartel, 1998; Chapple et al., 2003). Favorable interactions have been shown to occur between clays, liposomes, and RNAs (Hanczyc et al., 2003). And model experiments suggest that emergent, systems-level behavior could be present in fatty acid vesicles bearing ribozymes where the only function is RNA replication (Chen et al., 2004).

It is not always appreciated that in addition to metabolic functions, even very simple cells would need active regulatory mechanisms, particularly to temper stochastic effects resulting from small concentrations of constituents (McAdams and Arkin, 1999). Reassuringly, there are many natural examples of RNA aptamer-regulated mRNA translation in both prokaryotes (Barrick et al., 2004) and eukaryotes (reviewed in Kaempfer, 2003; also see Chapter 2.5). Likewise, *in vitro* selected "aptazymes" that transduce aptamer ligand binding to the control of

another activity indicate that RNAs capable of mediating responses to changing conditions can be readily isolated (Koizumi et al., 1999).

Finally, a simple RNA organism needs an efficient mechanism for catalyzing the replication of its biopolymeric structures. The class I ligase ribozyme has been engineered to polymerize 14 bases along an RNA template using ribonucleoside triphosphates as substrates (Johnston et al., 2001). However, the ribozyme itself is more than 200 nucleotides long. With such poor activity, molecules that large could never copy other polymerase structures, much less other functional sequences, before the polymerases themselves would chemically degrade. Perhaps polymerases made from more stable heteropolymers, such as DNA or threose nucleic acids (TNA) (reviewed in Eschenmoser, 1999, 2004), can be isolated, which, because of their better resistance to degradation, would have a better chance of exhibiting replicase behavior. Self-replicating ligase structures have been created (Paul and Joyce, 2002) though their capacity for open-ended evolution is limited because the "monomer" units are dozens of nucleotides long.

1.5
Recent Technological Developments and Future Directions

Breaker and colleagues (Koizumi et al., 1999) demonstrated a method for isolating aptazymes, or aptamer-controlled ribozymes, whereby an obligatory internal stem of a hammerhead ribozyme motif is replaced with a random sequence library (Fig. 1.6). Ligand binding to the random sequence domain induces a conformational shift, after which the ribozyme adopts a reaction-competent state. Aptazymes have been selected that are responsive to a wide variety of ligands, including small molecule metabolites (Seetharaman et al., 2001; Hesselberth et al., 2003) and peptides (Robertson et al., 2004). The relative ease with which aptazymes can be generated is related to the fact that many nucleic acid aptamers are relatively unstructured in the absence of ligand (Hermann and Patel, 2000). The binding loops tend to be rather floppy. And the flanking stems, if short enough, are energetically unstable until the loop becomes structured upon ligand binding.

The aptamer domain does not have to be explicitly selected as an aptazyme in order to function properly in that context. Active aptazyme constructs can be designed through simple secondary structure modeling. For example aptazyme constructs made with a pre-existing theophylline aptamer (Jenison et al., 1994) showed regulated group I intron splicing *in vivo* (Thompson et al., 2002).

Significantly, allosteric selections for aptamers do not require the ligand to have a linker or other chemical modification (Wilson and Szostak, 1999). Ordinarily, small molecule ligands have to be immobilized on a solid support to partition functional aptamer sequences away from inactive ones. This aspect of the method permits aptamers that completely enclose the ligand to be isolated, potentially increasing the maximum binding affinity for which an aptamer selection can be performed.

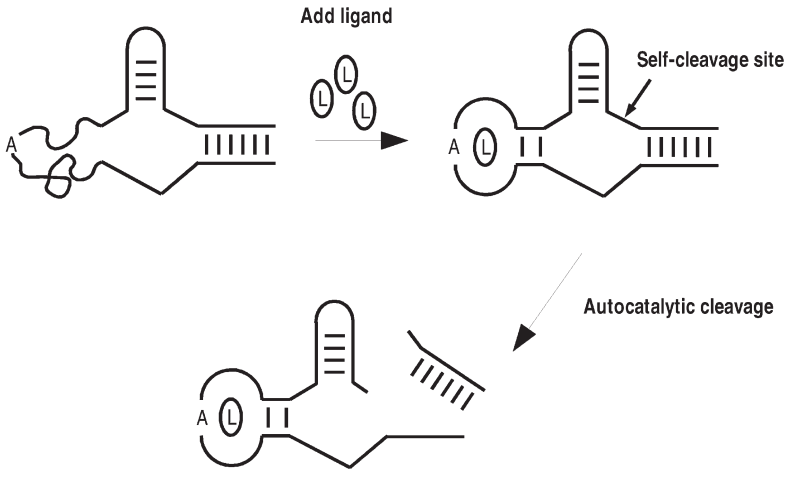


Fig. 1.6 A simple aptazyme. There are many different ways to couple aptamer ligand binding to the activation of a ribozyme (also see Chapters 9 and 16). This figure illustrates a simple self-cleaving aptazyme (after Koizumi et al., 1999). The structure consists of a self-cleaving hammerhead ribozyme with an internal stem replaced by an aptamer sequence ("A"). Upon binding the ligand ("L"), the aptamer becomes well-ordered, reconstituting the disrupted hammerhead stem. Once the stem forms, the ribozyme can adopt an active conformation and undergo self-cleavage.

When designing an *in vitro* selection protocol, it is important to make the activity needed to survive the selective step resemble the desired function as closely as possible. Most successful attempts to engineer multiple-turnover catalysts from random sequence libraries have actually begun with molecules isolated because they could perform a single-turnover reaction on a covalently attached substrate (for example see Bartel and Szostak, 1993; Seelig and Jaschke, 1999; reviewed in Griffiths and Tawfik, 2000). After the selection, the molecules must be re-engineered to function as true multiple-turnover enzymes (for example Ekland et al., 1995). In contrast, compartmentalizing each of the library molecules in its own reaction chamber allows for the direct selection of multiple-turnover catalysts (reviewed in Griffiths and Tawfik, 2000; Fig. 1.7).

In a typical protocol for selection by compartmentalization, water-in-oil emulsions are prepared by mixing mineral oil and a cocktail of detergents with a water phase containing DNA, protein polymerases, and buffer and substrates for transcription and/or translation. The emulsion compartments are stable over a range of temperature and buffer conditions. Catalysts can be selected on the basis of their ability to modify an exogenous substrate attached to a library DNA molecule enclosed in the same compartment (Tawfik and Griffiths, 1998). Fluorescence-activated cell sorting (FACS) can be used to enrich active molecules if the reaction can produce, directly or indirectly, a fluorescent signal. FACS has been employed to isolate active library sequences attached to beads (Sepp et al., 2002; Griffiths and Tawfik, 2003). Alternatively, water-in-oil compartments can be emulsified in

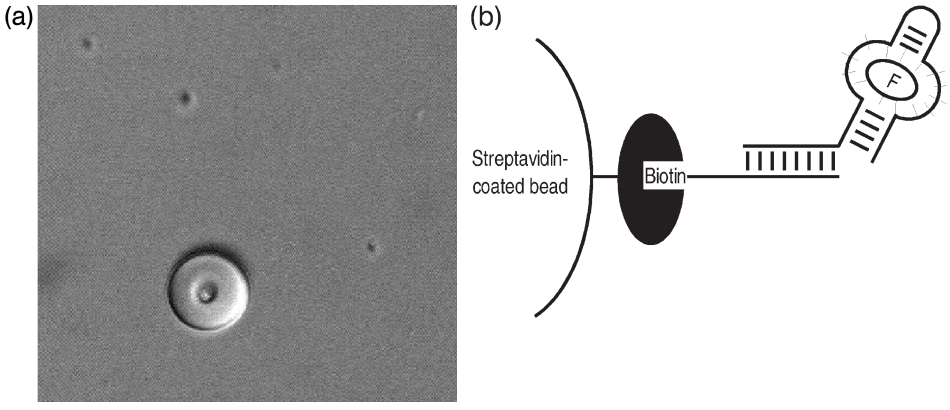


Fig. 1.7 Compartmentalized selections with emulsions, beads, and flow cytometry.
(a) Fluorescence from Oregon Green dye marks the aqueous droplet of a water-in-oil-in-water emulsion (also see Griffiths and Tawfik, 2003). The droplet in the center of the image contains RNA transcribed from a single DNA template (courtesy of Luptak and Szostak, unpublished). Water-in-oil droplets containing RNAs with a desired activity can be selected on the basis of a spectroscopic signal using flow cytometry. This approach also allows for the direct selection of aptamers or ribozymes that elicit a particular change in the emission spectrum of a fluorescent dye.
(b) In much the same way, flow cytometry can be used to sort microbeads displaying a spectroscopic signal. In the example depicted, RNA is immobilized on beads in an emulsified solution containing a fluorophore ("F"). After breaking the emulsions, beads displaying aptamers that induce changes in fluorescence can be sorted (after Luptak and Szostak, unpublished).

water to produce stable water-in-oil-in-water double emulsion compartments that can be sorted (Bernath et al., 2004). Liposomes also have properties that make them attractive for use as encapsulated reaction chambers (reviewed in Monnard, 2003). However, the emulsion techniques are much farther along in development because liposomes typically exhibit low encapsulation efficiency.

To produce reaction chambers 1 μm in diameter, approximately 1 mL of water-in-oil emulsion is needed for every 10^{12} library molecules (Tawfik and Griffiths, 1998) unless more than one library sequence is put into the same chamber. This limits efficient selections to a library size of about 10^{14} different molecules. Because FACS is a serial screening process, selections that employ it are further limited; 10^7 beads or compartments can be easily sorted but approaching even 10^{10} molecules would be nearly impossible. Bead panning (Coomber, 2002) is an alternative sorting process that may allow for higher throughput. However, until the sorting of larger libraries can be routinely accomplished, the best applications of compartmentalized selection are (1) selecting for simple activities that are very abundant in random sequence pools and (2) searching the sequence space surrounding pre-existing structures for variants with altered, or superior, activity.

Compartmentalized selection of protein enzymes has been used to improve a phosphotriesterase (Griffiths and Tawfik, 2003), alter the sequence specificity of

a methyltransferase (Cohen et al., 2004) and enhance the temperature stability and heparin resistance of Taq polymerase (Ghadessy et al., 2001). However, there is nothing that prevents compartmentalized selections from being used for the *in vitro* selection of functional nucleic acids.

Generating ribozymes with replicase activity remains a significant hurdle to the goal of creating simple chemical systems capable of Darwinian evolution (Szostak, 2001 et al., 2001; Luisi, 2002; Benner, 2003). In principle it should be possible to build replicases through iterations of *in vitro* selection and rational design. If there are class I ligase-based polymerase variants with improved affinity for the primer–template complex (Lawrence and Bartel, 2003), these could be isolated in a compartmentalized selection based directly on their ability to polymerize longer strands of RNA. Alternatively, compartmentalized selections for completely novel ribozymes with efficient and processive polymerase activity could be undertaken. In either case, these efforts might benefit from the use of substrates that are more highly activated than nucleotide triphosphates, or from the addition of other metal ions and cofactors (Szostak et al., 2001; for example Joyce et al., 1984, 1987).

Compartmentalization can also be used to isolate aptamers or ribozymes where a spectroscopic signal is the desired function, not just the readout. Two aptamers have been shown to elicit fortuitous changes in the spectroscopic properties of fluorescent small molecule ligands upon binding (Babendure et al., 2003, references therein). A FACS-based selection could be used to isolate aptamers that induce specific binding-dependent changes in quantum yield or spectroscopic resonance (Fig. 1.7) (Luptak and Szostak, unpublished results).

We briefly considered the issue of whether functional molecules can be built from oligonucleotides that are chemically distinct from RNA in Section 1.3. Functional nucleic acids that are more resistant to degradation than RNA have biotechnological applications. It is also possible that increasing the chemical diversity of the constituents used to build a library may improve the quality of the functional molecules that can be obtained (see Section 1.3). In Section 1.4 we alluded to the fact that while RNA is a good model for the first biopolymer, a wide range of possibilities should be explored.

Chemical alternatives to the ribose sugar backbone have been examined in some detail (reviewed in Eschenmoser, 1999, 2004). Among these, threose nucleic acid (TNA), with a four-carbon sugar backbone, is the most noteworthy. Because TNA is chemically simpler than RNA but retains many of its other properties, it is considered a plausible alternative to RNA as one of the earliest biopolymers (Eschenmoser, 2004). A DNA duplex containing a single TNA nucleotide maintains good helical geometry (Pallan et al., 2003), and single-stranded TNA base pairs with both RNA and DNA (Schoning et al., 2000).

Protein polymerases can incorporate TNA monomers by primer extension on DNA templates (Chaput and Szostak, 2003), and can also use TNA oligos to template DNA synthesis (Chaput et al., 2003). A system for selecting functional TNA structures from libraries of random molecules was recently devised (Fig. 1.8) (Ichida et al., 2005). Briefly, TNA sequences are displayed on DNA by primer-

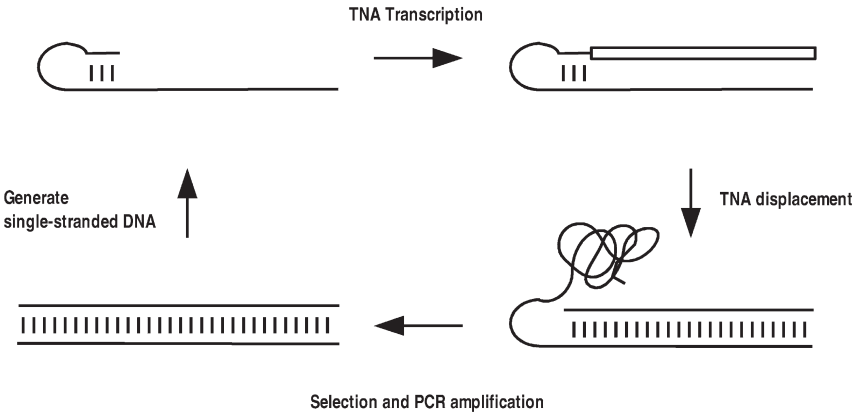


Fig. 1.8 *In vitro* selection for functional non-standard oligonucleotides. Any molecule whose synthesis can be templated on DNA is amenable to selection using DNA display. For instance, libraries of sequences with a four-carbon sugar backbone, threose nucleic acid (TNA) (reviewed in Eschenmoser, 1999, 2004) can be transcribed using the "therminator" DNA polymerase (Ichida et al., 2005). The TNA sequence is displaced from the template by extending a DNA primer in the opposite direction. Following the selection step, PCR amplifies the DNA template. Note that this method does not require enzymes to directly copy TNA because of the covalent linkage between the displayed sequence and its DNA template.

extending hairpins at the 3′ ends of library molecules. DNA sequences displaying functional TNA structures are amplified using PCR.

The TNA selection system is conceptually similar to nucleic acid display technologies for selecting proteins from random sequence libraries (Roberts and Szostak, 1997; Keefe and Szostak, 2001; Bertschinger and Neri, 2004; reviewed in Frankel et al., 2003) and directing the organic synthesis of diverse classes of molecules (Rosenbaum et al., 2003; Gartner et al., 2004; Kanan et al., 2004; Halpin and Harbury, 2004; reviewed in Calderone and Liu, 2004). Experimental results demonstrate that a large number of modified nucleobase-pairing systems can be constructed as long as rules for stacking, charge, geometry, and hydrogen bonding are followed (Geyer et al., 2003; Liu et al., 2003). Recent work has identified protein polymerases that can synthesize strands of other modified oligonucleotides (Ghadessy et al., 2004; Chelliserrykattil and Ellington, 2004; reviewed in Benner, 2004). These, or other novel enzymes, could be utilized to select functional molecules from any type of heteropolymer library displayed on DNA.

It will be interesting to see if functional activities, in addition to base-pairing ability, can be obtained from libraries of alternative oligonucleotides such as TNA. Experimentally determining the abundance and distribution of function in various oligonucleotide sequence spaces may shed light on why RNA, as opposed to TNA, or something else, is found in extant biology. By quantifying the number of functional sequences (Szostak, 2003) in sequence space or the diffi-

culty of specifying functional structures (Carothers et al., 2004), it will be possible to rigorously compare the functional capabilities of different kinds of oligonucleotides. Among the most chemically plausible prebiotic or early-biotic nucleic acids, RNA may turn out to be the most likely to exhibit functional activities. Alternatively, RNA may have become the evolutionary winner because of historically contingent events or contexts.

1.6 Conclusion

In vitro selection was developed by the Gold, Joyce and Szostak laboratories in the early 1990s to search nucleic acid sequence spaces for functional molecules. Since then, *in vitro* selection experiments have been performed by researchers around the world to isolate, optimize and characterize hundreds of different synthetic aptamers and catalysts.

In vitro selection has helped us understand the functional capabilities of oligonucleotides. Work to describe the distribution of functional structures in molecular sequence spaces continues to inform the design of sequence libraries with a greater likelihood of yielding highly active molecules. Study of the relationships between complexity and activity is teaching us how to isolate molecules of higher activity, and is also beginning to improve our comprehension of the origins of biochemistry. Heteropolymers would have been required at a very early stage in the history of life to carry out the necessary biochemical functions. While RNA may not have been the first biopolymer, it serves as a useful model because it possesses many of the qualities that the initial biopolymers would have needed. Recent work strongly suggests that the functional capabilities of other oligonucleotides can be examined in the near future.

New techniques and applications for *in vitro* selection continue to be developed. For example, encapsulated selections will enable direct enrichment for multiple turnover catalysts such as ribozymes with RNA polymerase activity, or for aptamers that elicit changes in the spectroscopic properties of the ligand. Once selections for functional structures made up of alternative nucleic acids such as TNA have been successful, it will be interesting to compare their functional properties with those of RNA and DNA. Is the number of functional structures in an oligonucleotide sequence space relatively constant for different types of molecules, or does it depend heavily on chemical composition? Are there general rules that describe how the number of functional structures or sequences varies with activity, or is every case unique? There are many fundamental questions about the origins of biochemical function and life on Earth that remain unanswered; some may even be unanswerable. We are hopeful that through the careful application of *in vitro* selection and other synthetic methods, we will come to understand the range of possible pathways that lead from chemical systems to biological ones.

Acknowledgments

We thank A. Luptak for sharing unpublished results and providing the image for Fig. 1.7a. We thank A. Luptak and J.K. Ichida for helpful discussions. Work in the authors' laboratory is supported by a grant from the National Institutes of Health (USA) (GM53936 to J.W. Szostak). J.M. Carothers was supported by a Graduate Fellowship from the National Science Foundation (USA). J.W. Szostak is an investigator of the Howard Hughes Medical Institute.

References

Adami, C. (2004). Information theory in molecular biology. *Phys Life Rev* 1, 3–22.

Adami, C., Cerf, N. J. (2000). Physical complexity of symbolic sequences. *Physica D (Amsterdam)* 37, 62–69.

Anet, F.A. (2004). The place of metabolism in the origin of life. *Curr Opin Chem Biol* 8, 654–659.

Babendure, J.R., Adams, S.R., Tsien, R.Y. (2003). Aptamers switch on fluorescence of triphenylmethane dyes. *J Am Chem Soc* 125, 14716–14717.

Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J.K., Breaker, R.R. (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci USA* 101, 6421–6426.

Bartel, D.P., Szostak, J.W. (1993). Isolation of a new ribozymes from a large pool of random sequences. *Science* 261, 1411–1418.

Battersby, T.R., Ang, D.N., Burgstaller, P., Jurczyk, S.C., Bowser, M.T., Buchanan, D.D., Kennedy, R.T., Benner, S.A. (1999). Quantitative analysis of receptors for adenosine nucleotides obtained via *in vitro* selection from a library incorporating a cationic nucleotide analog. *J Am Chem Soc* 121, 9781–9789.

Been, M.D., Cech, T.R. (1988). RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the *Tetrahymena* ribozyme. *Science* 238, 1412–1416.

Benner, S.A. (2003). Synthetic biology: Act natural. *Nature* 421, 118.

Benner, S.A. (2004). Understanding nucleic acids using synthetic chemistry. *Acc Chem Res* 37, 784–797.

Benner, S.A., Burgstaller, P., Battersby, T.R., Jurczyk, S (1999). Did the RNA world exploit an expanded genetic alphabet? In: *The RNA World*, 2nd edn, Gesteland, R.F., Cech, T.R., Atkins, J.F., eds. New York: Cold Spring Harbor Press, pp. 163–181.

Bergman, N.H., Johnston, W.K., Bartel, D.P. (2000). Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* 39, 3115–3123.

Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A.D., Magdassi, S., Tawfik, D.S. (2004). *In vitro* compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Anal Biochem* 325, 151–157.

Bertschinger, J., Neri, D. (2004). Covalent DNA display as a novel tool for directed evolution of proteins *in vitro*. *Protein Eng Des Sel* 17, 699–707.

Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., Toole, J.J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355, 564–566.

Breaker, R.R., Banerji, A., Joyce, G.F. (1994). Continuous *in vitro* evolution of bacteriophage RNA polymerase promoters. *Biochemistry* 33, 11980–11986.

Breslow, R. (2001). Biomimetic selectivity. *Chem Rec* 1, 3–11.

Brown, R.S., Hingerty, B.E., Dewan, J.C., Klug, A. (1983). Pb(II)-catalysed cleavage of the sugar-phosphate backbone of yeast $tRNA^{Phe}$ – implications for lead toxicity and self-splicing RNA. *Nature* 303, 543–546.

Burgstaller, P., Famulok, M. (1994). Isolation of RNA aptamers for biological cofactors by

in vitro selection. *Angew Chem Int Ed Engl* 33, 1084–1087.

Burke, D.H., Gold, L. (1997). RNA aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX. *Nucleic Acids Res* 25, 2020–2024.

Calderone, C.T., Liu, D.R. (2004). Nucleic-acid-templated synthesis as a model system for ancient translation. *Curr Opin Chem Biol* 8, 645–653.

Cane, D.E., Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem Biol* 6, R319–325.

Carothers, J.M., Oestreich, S.C., Davis, J.H., Szostak, J.W. (2004). Informational complexity and functional activity of RNA structures. *J Am Chem Soc* 126, 5130–5135.

Cech, T.R. (1986). A model for the RNA-catalyzed replication of RNA. *Proc Natl Acad Sci USA* 83, 4360–4363.

Cech, T.R. (2002). Ribozymes, the first 20 years. *Biochem Soc Trans* 30, 1162–1166.

Chapple, K.E., Bartel, D.P., Unrau, P.J. (2003). Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *RNA* 9, 1208–1220.

Chaput, J.C., Szostak, J.W. (2003). TNA synthesis by DNA polymerases. *J Am Chem Soc* 125, 9274–9275.

Chaput, J.C., Ichida, J.K., Szostak, J.W. (2003). DNA polymerase-mediated DNA synthesis on a TNA template. *J Am Chem Soc* 125, 856–857.

Check, E. (2002). Venter aims for maximum impact with minimal genome. *Nature* 420, 350.

Chelliserrykattil, J., Ellington, A.D. (2004). Evolution of a T7 RNA polymerase variant that transcribes 2'-O-methyl RNA. *Nat Biotechnol* 22, 1155–1160.

Chen, I.A., Roberts, R.W., Szostak, J.W. (2004). The emergence of competition between model protocells. *Science* 305, 1474–1476.

Cohen, H.M., Tawfik, D.S., Griffiths, A.D. (2004). Altering the sequence specificity of HaeIII methyltransferase by directed evolution using *in vitro* compartmentalization. *Protein Eng Des Sel* 17, 3–11.

Coomber, D.W. (2002). Panning of antibody phage-display libraries. Standard protocols. *Methods Mol Biol* 178, 133–145.

Cordova, A. (2004). The direct catalytic asymmetric mannich reaction. *Acc Chem Res* 37, 102–112.

Davis, J.D., Szostak, J.W. (2002). Isolation of high-affinity GTP aptamers from partially-structured RNA libraries. *Proc Natl Acad Sci USA* 99, 11616–11621.

Dieckmann, T., Suzuki, E., Nakamura, G.K., Feigon, J. (1996). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA* 2, 628–640.

Doudna, J.A., Cech, T.R. (2002). The chemical repertoire of natural ribozymes. *Nature* 418, 222–228.

Doudna, J.A., Szostak, J.W. (1989). RNA-catalysed synthesis of complementary-strand RNA. *Nature* 339, 519–522.

Eaton, B.E., Gold, L., Zichi, D.A. (1995) Let's get specific: the relationship between specificity and affinity. *Chem Biol* 2, 633–638.

Ekland, E.H., Bartel, D.P. (1995). The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res* 23, 3231–3238.

Ekland, E.H., Bartel, D.P. (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* 382, 373–376.

Ekland, E.H., Szostak, J.W., and Bartel, D.P. (1995). Structurally complex and highly active RNA ligases derive from random RNA sequences. *Science* 269, 364–370.

Eliceiri, G.L. (1999). Small nucleolar RNAs. *Cell Mol Life Sci* 56, 22–31.

Ellington, A.D., Szostak, J.W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* 346, 818–822.

Ellington, A.D., Szostak, J.W. (1992). Selection *in vitro* of single-stranded DNA molecules that fold into specific ligand binding structures. *Nature* 355,850–852.

Eschenmoser, A. (1999). Chemical etiology of nucleic acid structure. *Science* 284, 2118–2124.

Eschenmoser, A. (2004). The TNA-family of nucleic acid systems: properties and prospects. *Orig Life Evol Biosph* 34, 277–306.

Frankel, A., Li S., Starck, S.R., Roberts, R.W. (2003). Unnatural RNA display libraries. *Curr Opin Struct Biol* 13, 506–512.

Gardiner, K., Pace, N.R. (1980). RNase P of *Bacillus subtilis* has a RNA component. *J Biol Chem* 255, 7507–7509.

Gartner, Z.J., Tse, B.N., Grubina, R., Doyon, J.B., Snyder, T.M., Liu, D.R. (2004). DNA-templated organic synthesis and selection of a library of macrocycles. *Science* 305, 1601–1605.

Geyer, C.R., Battersby, T.R., Benner, S.A. (2003). Nucleobase pairing in expanded Watson-Crick-like genetic information systems. *Structure (Camb)* 11, 1485–1498.

Ghadessy, F.J., Ong, J.L., Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci USA* 98, 4552–4557.

Ghadessy, F.J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., Vaisman, A., Woodgate, R., Holliger, P. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat Biotechnol* 22, 755–759.

Gilbert, W. (1986). The RNA World. *Nature* 319, 618.

Glasner, M.E., Bergman, N.H., Bartel, D.P. (2002). Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry* 41, 8103–8112.

Godfrey-Smith, P. (2002). On genetic information and genetic coding. In: *The Scope of Logic, Methodology, and the Philosophy of Science*, Vol. II. P. Gardenfors, J. Wolenski, K. Kajania-Placek, eds. Dordrecht: Kluwer, pp. 387–400.

Green, R., Ellington, A.D., Szostak, J.W. (1990). *In vitro* genetic analysis of the *Tetrahymena* self-splicing intron. *Nature* 347, 406–408.

Griffiths, A.D., Tawfik, D.S. (2000). Man-made enzymes – from design to *in vitro* compartmentalisation. *Curr Opin Biotechnol* 11, 338–353.

Griffiths, A.D., Tawfik, D.S. (2003). Directed evolution of an extremely fast phosphotriesterase by *in vitro* compartmentalization. *EMBO J* 22, 24–35.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849–857.

Halpin, D.R., Harbury, P.B. (2004). DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biol* 2, E174.

Halpin, D.R., Lee, J.A., Wrenn, S.J., Harbury, P.B. (2004). DNA display III. Solid-phase organic synthesis on unprotected DNA. *PLoS Biol* 2, E175.

Hanczyc, M. M., Dorit, R. L. (2000). Replicability and recurrence in the experimental evolution of a group I ribozyme. *Mol Biol Evol* 17, 1050–1060.

Hanczyc, M.M., Szostak, J.W. (2004). Replicating vesicles as models of primitive cell growth and division. *Curr Opin Chem Biol* 8, 660–664.

Hanczyc, M.M., Fujikawa, S.M., Szostak, J.W. (2003). Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* 302, 618–622.

Held, D.M., Greathouse, S.T., Agrawal, A., Burke, D.H. (2003). Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers. *J Mol Evol* 57, 299–308.

Hermann, T., Patel, D.J. (2000). Adaptive recognition by nucleic acid aptamers. *Science* 287, 820–825.

Hesselberth, J.R., Robertson, M.P., Knudsen, S.M., Ellington, A.D. (2003). Simultaneous detection of diverse analytes with an aptazyme ligase array. *Anal Biochem* 312, 106–112.

Huang, Z., Szostak, J.W. (2003). Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA* 9, 1456–1463.

Huizenga, D.E., Szostak, J.W. (1995). A DNA aptamer that binds adenosine and ATP. *Biochemistry* 34, 656–665.

Ichida, J.K., Zou, K., Horhota, A., Yu, B., McLaughlin, L.W., Szostak, J.W. (2005). An in vitro selection system for TNA. *J Am Chem Soc* 127, 2802–2803.

Jenison, R. D., Gill, S. C., Pardi, A., Polisky, B. (1994). High-resolution molecular discrimination by RNA. *Science* 263, 1425–1429.

Jiang, F., Kumar, R.A, Jones, R.A., Patel, D.J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature* 382, 183–186.

Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., Bartel, D.P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292, 1319–1325.

Joyce, G.F. (1989). Amplification, mutation, and selection of catalytic RNA. *Gene* 82, 83–87.

Joyce, G.F. (2002). The antiquity of RNA-based evolution. *Nature* 418, 214–221.

Joyce, G.F., Inoue, T., Orgel, L.E. (1984). Non-enzymatic template-directed synthesis on RNA random copolymers. Poly(C, U) templates. *J Mol Biol* 176, 279–306.

Joyce, G.F., Schwartz, A.W., Miller, S., Orgel, L.E. (1987). The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc Natl Acad Sci USA* 84, 4398–4402.

Kaempfer, R. (2003). RNA sensors: novel regulators of gene expression. *EMBO Rep* 4, 1043–1047.

Kanan, M.W., Rozenman, M.M., Sakurai, K., Snyder, T.M., Liu, D.R. (2004). Reaction discovery enabled by DNA-templated synthesis and *in vitro* selection. *Nature* 431, 545–549.

Kauffman, S. (1986). Autocatalytic sets of proteins. *J Theor Biol* 119, 1–24.

Keefe, A.D., Szostak, J.W. (2001). Functional proteins from a random-sequence library. *Nature* 410, 715–718.

Knight, R., Yarus, M. (2003). Finding specific RNA motifs: function in a zeptomole world? *RNA* 9, 218–230.

Koizumi, M., Soukup, G.A., Kerr, J.N., Breaker, R.R. (1999). Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Biol* 6, 1062–1071.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31, 147–157.

Lauhon, C.T., Szostak, J.W. (1995). RNA aptamers that bind flavin and nicotinamide redox cofactors. *J Am Chem Soc* 117, 1246–1257.

Lawrence, M.S., Bartel, D.P. (2003). Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry* 42, 8748–8755.

Lehman, N., Joyce, G.F. (1993). Evolution *in vitro*: analysis of a lineage of ribozymes. *Curr Biol* 3, 723–34.

Lehman, N., Donne, M.D., West, M., Dewey, T.G. (2000). The genotypic landscape during *in vitro* evolution of a catalytic RNA: implications for phenotypic buffering. *J Mol Evol* 50, 481–490.

Li, Y., Sen, D. (1997). Toward an efficient DNAzyme. *Biochemistry* 36, 5589–5599.

Lilley, D.M. (2003). The origins of RNA catalysis in ribozymes. *Trends Biochem Sci* 28, 495–501.

Liu, H., Gao, J., Lynch, S.R., Saito, Y.D., Maynard, L., Kool, E.T. (2003). A four-base paired genetic helix with expanded size. *Science* 302, 868–871.

Lorsch, J.R., Szostak, J.W. (1996). Chance and necessity in the selection of nucleic acid catalysts. *Acc Chem Res* 29, 103–110.

Luisi, P.L. (2002). Toward the engineering of minimal living cells. *Anat Rec* 268, 208–214.

Mandal, M., Breaker, R.R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* 11, 29–35.

Martin, W., Russell, M.J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* 358, 59–83.

Maynard Smith, J. (1970). Natural selection and the concept of a protein space. *Nature* 225, 563–564.

McAdams, H.H., Arkin, A. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* 15, 65–69.

McKay, C.P. (2004). What is life – and how do we search for it in other worlds? *PLoS Biol* 2, e302.

Mojzsis, S.J., Krishnamurthy, R., Arrhenius, G. (1999). Before RNA and after: Geophysical and geochemical constraints on molecular evolution. In: *The RNA World*, 2nd edn, Gesteland, R.F., Cech, T.R., Atkins, J.F., eds. New York: Cold Spring Harbor Press, pp. 1–47.

Monnard, P.A. (2003). Liposome-entrapped polymerases as models for microscale/nanoscale bioreactors. *J Membr Biol* 191, 87–97.

Monnard, P.A., Deamer, D.W. (2002). Membrane self-assembly processes: Steps toward the first cellular life. *Anat Rec* 268, 196–207.

Okumoto, Y., Tanabe, Y., Sugimoto, N. (2003). Factors that contribute to efficient catalytic activity of a small Ca^{2+}-dependent deoxyribozyme in relation to its RNA cleavage function. *Biochemistry* 42, 2158–2165.

Orgel, L.E. (1986). RNA catalysis and the origins of life. *J Theor Biol* 123, 127–149.

Pallan, P.S., Wilds, C.J., Wawrzak, Z., Krishnamurthy, R., Eschenmoser, A., Egli, M. (2003). Why does TNA cross-pair more strongly with RNA than with DNA? an answer from X-ray analysis. *Angew Chem Int Ed Engl* 42, 5893–5895.

Pan, T., Uhlenbeck, O.C. (1992). *In vitro* selection of RNAs that undergo autolytic cleavage with Pb^{2+}. *Biochemistry* 31, 3887–3895.

Paul, N., Joyce, G.F. (2002). Inaugural article: a self-replicating ligase ribozyme. *Proc Natl Acad Sci USA* 99, 12733–12740.

Reader, J.S., Joyce, G.F. (2002). A ribozyme composed of only two different nucleotides. *Nature* 420, 841–844.

Reidys, C., Forst, C. V., Schuster, P. (2001). Replication and mutation on neutral networks. *Bull Math Biol* 63, 57–94.

Ricardo, A., Carrigan, M.A., Olcott, A.N., Benner, S.A. (2004). Borate minerals stabilize ribose. *Science* 303, 196.

Riley, C.A., Lehman, N. (2003). Expanded divalent metal-ion tolerance of evolved ligase ribozymes. *Biochimie* 85, 683–689.

Roberts, R.W., Szostak, J.W. (1997). RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc Natl Acad Sci USA* 94, 12297–12302.

Robertson, D.L., Joyce, G.F. (1990). Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* 344, 467–468.

Robertson, M.P., Ellington, A.D. (1997). Ribozymes: red in tooth and claw. *Curr Biol* 7, R376–R379.

Robertson, M.P., Knudsen, S.M., Ellington, A.D. (2004). *In vitro* selection of ribozymes dependent on peptides for activity. *RNA* 10, 114–127.

Rogers, J., Joyce, G.F. (1999). A ribozyme that lacks cytidine. *Nature* 402, 323–325.

Rosenbaum, D.M., and Liu, D.R. (2003). Efficient and Sequence-Specific DNA-Templated Polymerization of Peptide Nucleic Acid Aldehydes. *J Am Chem Soc* 125, 13924–13925.

Roth, A., Breaker, R.R. (1998). An amino acid as a cofactor for a catalytic polynucleotide. *Proc Natl Acad Sci USA* 95, 6027–6031.

Ruiz-Mirazo, K., Pereto, J., Moreno, A. (2004). A universal definition of life: autonomy and open-ended evolution. *Orig Life Evol Biosph* 34, 323–346.

Sabeti, P.C., Unrau, P.J., Bartel, D.P. (1997). Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem Biol* 4, 767–774.

Saffhill, R., Schneider-Bernloehr, H., Orgel, L.E., Spiegelman, S. (1970). *In vitro* selection of bacteriophage Q-beta ribonucleic acid variants resistant to ethidium bromide. *J Mol Biol* 51, 531–539.

Salehi-Ashtiani, K., Szostak, J.W. (2001). In vitro evolution suggests multiple origins for the hammerhead ribozyme. *Nature* 414, 82–84.

Sassanfar, M., Szostak, J.W. (1993). An RNA motif that binds ATP. *Nature* 364, 550–553.

Sazani, P.L., Larralde, R., Szostak, J.W. (2004). A small aptamer with strong and specific recognition of the triphosphate of ATP. *J Am Chem Soc* 126, 8370.

Schneider, F.W. (1985). Periodic perturbations of chemical oscillators: experiments. *Annu Rev Phys Chem* 36, 347–378.

Schneider, T. D., Stormo, G. D., Gold, L., and Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. *J Mol Biol* 188, 415–431.

Schoning, K., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R., Eschenmoser, A. (2000). Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3'→2') oligonucleotide system. *Science* 290, 1347–1351.

Schultes, E.A., Bartel, D.P. (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* 289, 448–452.

Seelig, B., Jaschke, A. (1999). A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol* 6, 167–76.

Seetharaman, S., Zivarts, M., Sudarsan, N., Breaker, R.R. (2001). Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat Biotechnol* 19, 336–341.

Sepp, A., Tawfik, D.S., Griffiths, A.D. (2002). Microbead display by *in vitro* compartmentalisation: selection for binding using flow cytometry. *FEBS Lett* 532, 455–458.

Sidorov, A.V., Grasby, J.A., Williams, DM. (2004). Sequence-specific cleavage of RNA in the absence of divalent metal ions by a DNAzyme incorporating imidazolyl and

amino functionalities. *Nucleic Acids Res* 32, 1591–1601.

Silverman, S.K. (2004). Deoxyribozymes: DNA catalysts for bioorganic chemistry. *Org Biomol Chem* 2, 2701–2706.

Simoneit, B.R. (2002). Molecular indicators (biomarkers) of past life. *Anat Rec* 268, 186–195.

Smith, H.O., Hutchison, C.A., Pfannkoch, C., Venter, C.J. (2003). Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci USA* 100, 15440–15445.

Steitz, T.A., Moore, P.B. (2003). RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* 28, 411–418.

Stemmer, W.P. (1994). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389–391.

Stubbe, J. (2000). Ribonucleotide reductases: the link between an RNA and a DNA world? *Curr Opin Struct Biol* 10, 731–736.

Szostak, J.W. (1988). Structure and activity of ribozymes. In: *Redesigning the Molecules of Life*. S.A. Benner, ed. Heidelberg: Springer-Verlag, pp. 87–114.

Szostak, J.W. (2003). Functional information: molecular messages. *Nature* 423, 689.

Szostak, J.W., Bartel, D.P., Luisi, L. (2001). Synthesizing life. *Nature* 409, 387–390.

Tawfik, D.S., Griffiths, A.D. (1998). Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 16, 652–656.

Thompson, K.M., Syrett, H.A., Knudsen, S.M., Ellington, A.D. (2002). Group I aptazymes as genetic regulatory switches. *BMC Biotechnol* 2, 21.

Tsukiji, S., Pattnaik, S.B., Suga, H. (2004). Reduction of an aldehyde by a NADH/Zn^{2+}-dependent redox active ribozyme. *J Am Chem Soc* 126, 5044–5045.

Tuerk, C., Gold, L. (1990). Systematic evolution of ligands by exponential enrichment. *Science* 249, 505–510.

Unrau, P.J., Bartel, D.P. (1998). RNA-catalysed nucleotide synthesis. *Nature* 395, 260–263.

Urbach, J.M. (2000). Factors that affect the outcome of *in vitro* selection. Harvard University doctoral thesis.

Vaish, N.K., Larralde, R., Fraley, A.W., Szostak, J.W., McLaughlin, L.W. (2003). A novel, modification-dependent ATP-binding aptamer selected from an RNA library incorporating a cationic functionality. *Biochemistry* 42, 8842–8851.

van Hest, J.C., Tirrell, D.A. (2001). Protein-based materials, toward a new level of structural control. *Chem Commun (Camb)* 19, 1897–1904.

Waugh, D.S., Green, C.J., Pace, N.R. (1989). The design and catalytic properties of a simplified ribonuclease P RNA. *Science* 244, 1569–1571.

Wilson, D.S., Szostak, J.W. (1999). *In vitro* selection of functional nucleic acids. *Annu Rev Biochem* 68, 611–647.

Wright, M.C., Joyce, G.F. (1997). Continuous *in vitro* evolution of catalytic function. *Science* 276, 614–617.

Yang, Y., Kochoyan, M., Burgstaller, P, Westhof, E., Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* 272, 1343–1347.

Zarrinpar, A., Park, S-H., Lim, W.A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 426, 676–680.

Zhao, H., Arnold, F.H. (1997). Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res* 25, 1307–1308.

Zinnen, S.P., Domenico, K., Wilson, M., Dickinson, B.A., Beaudry, A., Mokler, V., Daniher, A.T., Burgin, A., Beigelman, L. (2002). Selection, design, and characterization of a new potentially therapeutic ribozyme. *RNA* 8, 214–228.