1. Introduction

In vitro selection experiments have yielded a number of diverse aptamers. These RNA molecules bind tightly and with high specificity to their targets, but their binding mechanisms remain largely undetermined. Understanding the details of the binding process of RNA aptamers to their target molecules is however crucial for being able to define a general relationship between the structure of a molecule and its biological activity.

1.1 RNA - Functions and Properties

The RNA molecule has a pervasive role in contemporary biology, especially with regard to the most fundamental and highly conserved cellular processes. It is involved as a primer in DNA replication, as a messenger and as an adaptor (tRNA) in transferring genetic information to the translation machinery, and as a catalyst it is part of the ribosome. RNA instructs the processing of precursor messenger RNAs during splicing and editing and mediates numerous other transactions of RNA and proteins in the cell. Catalytic RNAs (ribozymes) act in RNA processing events and the replication of viral genomes. Individual ribonucleotides serve as important signaling molecules and as energy carriers, and their coenzyme derivatives participate in most of the reactions of central metabolism.

Despite containing only four different chemical subunits, RNA folds into a variety of complex tertiary structures (Ferre-d'Amare and Doudna, 1999; Gold and Eddy, 1993), similar to structured proteins. RNA therefore interacts specifically with other molecules and catalyses a broad range of chemical transformations (see reviews by Doudna and Cech, 2002, and by Joyce, 2002).

In 1998, Fire and Mello showed that double-stranded RNA could play an important role in regulating gene expression in eukaryotes (Fire et al., 1998; McManus and Sharp, 2002). Based on the silencing of specific genes by double-stranded RNA

(dsRNA), this technology was called RNA interference (RNAi) (for a review see Reinhart and Bartel, 2002).

The first example of an RNA molecule that forms a catalytically active site for a series of precise biochemical reactions was reported over 20 years ago: the self-splicing pre-ribosomal RNA of the ciliate *Tetrahymena*. The word 'ribozyme' was coined for a catalytic RNA molecule (Kruger et al., 1982). The following year catalytic activity was discovered in the RNA component of a ribonucleoprotein enzyme, ribonuclease (RNase) P, providing the first example of a multiple-turnover enzyme using RNA-based catalysis (Guerrier-Takada et al., 1983). The list of naturally occurring RNA catalysts has continued to expand through the identification of further self-splicing (Peebles et al. 1986) and also self-cleaving (Prody et al., 1986; Sharmeen et al., 1988; Saville et al., 1990) ribozymes. In addition, it has been shown that the catalytic components of the ribosome are made of RNA (Sharp, 1991; Noller et al., 1992; for reviews see Breaker, 1997b, and Doudna and Cech, 2002).

1.2 The RNA World Hypothesis

RNA encodes genomic information and is also able to function as a catalyst. In contemporary biology, DNA contains most genomic information and proteins are the major catalysts, but it is assumed that there was a time when life on earth was based on RNA rather than on DNA and protein (Woese, 1967; Crick, 1968; Orgel, 1968; Pace and Marsh, 1985; Sharp, 1985; Cech, 1986; Gilbert, 1986; Lewin, 1986; Joyce, 1991). This era is referred to as the “RNA world” (for a review see Joyce, 2002).

The messenger, transfer and ribosomal RNA molecules that exist in all known organisms direct the assembly of specific polypeptide sequences, instructed by corresponding RNA sequences. The earliest cells were likely to have relied on RNA for information storage and catalysis (Gilbert, 1986), with DNA and proteins arising as later specializations.

It is not known whether the invention of protein synthesis preceded or followed

that of DNA genomes. The primary advantage of DNA over RNA as a genetic material is the greater chemical stability of DNA, allowing much larger genomes based on DNA. Proteins may also require more genetic information than can be maintained by RNA, but *in vitro*-evolved ribozymes have achieved many of the tasks that proteins are able to perform. Additional catalytic RNAs are likely to be found in biology and undoubtedly many more will be discovered through *in vitro* selection. The construction of artificial RNA-based life from synthetic oligonucleotides is therefore a distinct possibility (Szostak et al., 2001).

1.3 RNA Structure, Fitness and Information Content

The primary structure of functional RNA molecules determines their structure by driving their self-assembly. This process has large negative free energy values, and the native structure represents a stable free-energy minimum within a landscape of almost infinite possible folded configurations (Wolynes et al., 1995). The movement into highly ordered form is therefore largely spontaneous.

The base-pairing pattern of RNA secondary structure provides both a geometric and thermodynamic scaffold for the tertiary structure of the molecule, thus connecting the secondary structure to functional properties of the tertiary structure. Genetic changes are therefore phenotypic changes in RNA (Ancel and Fontana, 2000).

To describe the relationship of aptamer activity and abundance in a pool of sequences of known length, design elements and diversity, information content can serve as a parameter. According to classical information theory, information content is defined as the amount of information required to specify a unique sequence or structure (Shannon and Weaver, 1963; Schneider et al., 1986; Adami and Cerf, 2000) and can also be converted into a parameter describing the probability of isolating a particular aptamer in sequence space as well as aptamer tolerance to mutations.

Sequence space, the ensemble of all possible sequences combined, is amorphous in the absence of functional demand, but can be mapped onto a fitness landscape where

every point is assigned a fitness value for a particular environment (Eigen et al., 1989; Kauffman and Macready, 1995). Within the fitness landscape, peaks represent identities that perform the function efficiently. Populations, natural or artificial, cover connected areas in sequence space. Sequences can evolve within a sub-region around a local peak, but passage to more distant peaks is prohibited by the demands of fitness on the phenotype and only feasible if regions of critically low fitness can be avoided.

Nucleic acid sequence space is populated by functional structures. Due to the massive combinatorial power of biological sequence space, even moderate RNA length however generates vast possibilities and adds to the difficulty of isolating these functional molecules. The number of independent isolates from a sequence library representing a given motif thus provides a qualitative measure of the information content of that motif: If a specific motif can be formed by a large number of independent isolates, it exhibits high abundance in the library, presumably due to its low information content. If however only a few independent sequences can form a given motif, its low abundance most likely correlates with its high information content.

1.4 *In Vitro* Selection

Even though sequence space is populated by many functional molecules, these are difficult to isolate. Since RNA can present a phenotype directly while acting synonymously as a genotype for the replication and amplification of that phenotype, schemes for rapid *in vitro* selection could be devised. *In vitro* selection or SELEX (Systematic Evolution of Ligands by Exponential Enrichment) (Ellington and Szostak, 1990; Robertson and Joyce, 1990; Tuerk and Gold, 1990) thus provides a route to rapidly isolate RNA sequences with specific functional properties such as binding or catalytic activity by exploiting the union of genotype and phenotype in RNA in a single molecule. A partial or complete sequence space for a given polymeric domain can be generated as a random sequence library. In order to isolate functional RNA molecules, some discriminatory principle is then coupled with enzymatic amplification to enrich for

individuals exhibiting desired fitness. Variation is introduced into the population either by artificially increased mutation rates or by partial randomization of RNA sequences. This process can be described as an adaptive walk, analogous to biological evolution, which traverses a region of nucleic acid sequence space, locating points of information, or peaks in the fitness landscape, via selection. A typical selection is shown in Fig. 1:

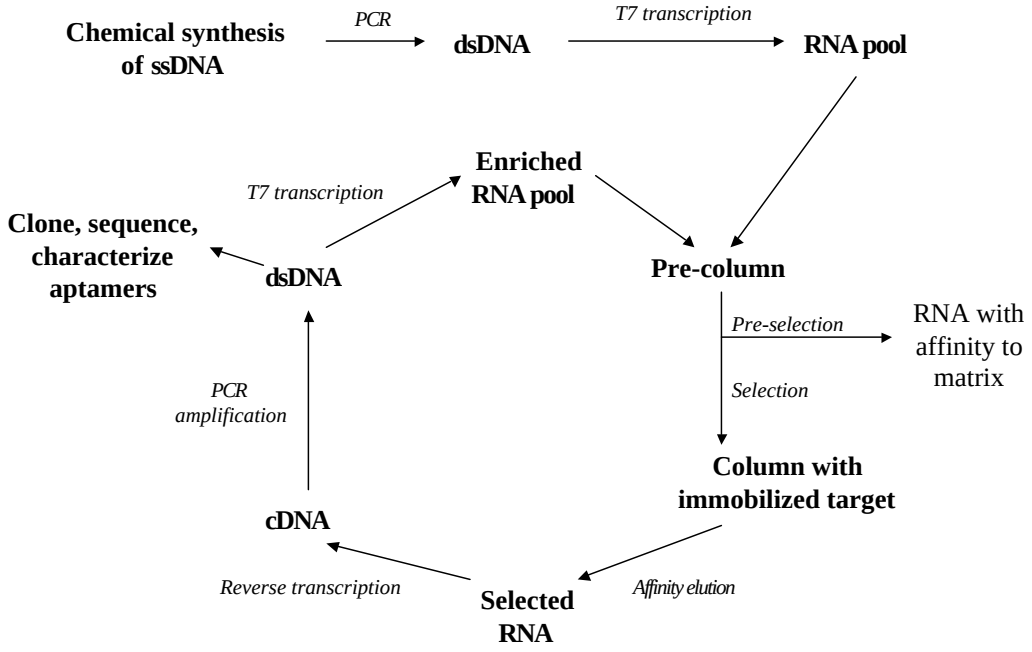

Fig. 1: *In Vitro* Selection Cycle. A library of chemically synthesized DNA flanked by both 3' and 5' defined constant regions (see pool sequences in the Materials chapter) with a large number of unique sequences (>10^{15}) is transcribed *in vitro*. The resulting RNA library is subjected to some selection pressure that enriches the functional sequences. The retained RNAs are eluted, reverse transcribed, amplified by PCR, transcribed, and then the entire cycle is repeated. The cycles continue until a significant fraction of the enriched library contains functional aptamers or ribozymes.

Little is known *a priori* about the abundance and distribution of active molecules in sequence space. Searches are therefore performed with random-sequence libraries that sample molecules scattered throughout the entire space. Since only approximately 10^{15} library molecules of length L can be practically synthesized, the coverage of sequence

space when sampling it can range from exhaustive (L≤25) to vanishingly sparse (L>25), depending on the length of the random region of the library. Therefore, nucleic acid libraries typically used in selection experiments (L>26) are incomplete and contain only a fraction of all possible sequences or presentations of functional motifs.

Gaining a quantitative understanding of the distribution of functional sequences in sequence space would greatly contribute to the effectiveness of *in vitro* selections if this knowledge can be translated into improving the design of RNA pools. Insights into the abundance, variability and complexity of specific solutions that are capable of solving a distinct biochemical problem would represent a fundamental advancement for the field attempting to elucidate issues pertaining to the origin of life.

1.5 Brief History of *in Vitro* Selection

The first to exploit the multiple properties of RNA was Sol Spiegelman, whose work in the 1960s with the RNA bacteriophage Qß showed that Darwinian selection could operate in a cell-free system (reviewed in Spiegelman, 1971).

In 1988 and 1989, Szostak and Joyce respectively described a scheme of evolution that can be used to isolate new catalytic ribonucleic acids (Szostak, 1988; Joyce, 1989a). By the 1990s it was possible to generate immense sequence diversity by chemically synthesizing DNA in which some regions consisted of completely random sequence. The isolation of reverse transcriptase and the invention of polymerase chain reaction (PCR) made it easy to replicate almost any nucleic acid sequence *in vitro* (Saiki et al., 1985) without the constraints of the Qß system. When combined, these advances enabled the development of *in vitro* selection and directed evolution to search sequence space for new functional RNA sequences (Ellington and Szostak, 1990; Green et al., 1990; Robertson and Joyce, 1990; Tuerk and Gold, 1990).

Considerable progress has been achieved in this field since the first examples of the use of combinatorial nucleic acid libraries for the *in vitro* selection of specific ligand-

binding RNAs in 1990. Nucleic acid aptamers for more than a hundred different targets have been described.

The range of applications of this technology extends from basic research to the identification of novel diagnostic and therapeutic reagents (Sullenger and Gilboa, 2002; Recent Patents in Aptamer Technology, *Nat. Biotechnol.* **20**, 403 (2002); Gold et al., 1997; Tian et al., 1995; Davis et al., 1996; Drolet et al., 1996), including the evaluation of binding sites between proteins and nucleic acids, the study of structure, function and substrate specificity of ribozymes and *de novo* isolation and evaluation of nucleic acids that bind to small biological and abiotic organic molecules (Famulok and Szostak, 1993). Aptamers show high affinity and specificity for their target molecule, comparable to monoclonal antibodies. Applications such as diagnostics and therapeutics (Gold, 1995) are however limited due to the instability of RNA molecules *in vivo*. Therefore various attempts to stabilize the aptamers have also been the subject of research (Eaton and Pieken, 1995; Klußmann et al., 1996; Nolte et al., 1996).

The specificity of molecular recognition combined with the simplicity by which aptamers can be obtained, engineered, and chemically modified make these molecules very attractive as drugs or tools in biotechnology and diagnostics (Famulok and Jenne, 1998). SELEX technology could thus play a role in affinity chromatography, medical diagnostics, and medical imaging *in vivo* (Hartig et al., 2002; Gold et al., 1995 and 1997; for reviews see Jaeger, 1997; Breaker, 1997b).

1.6 Aptamers and Catalytic RNAs Generated by *In Vitro* Selection

Sequence space has been searched (Smith, 1970) for functional RNA, DNA, and protein molecules (Lamla and Erdmann, 2003; Schuster, 2002; Dower and Mattheakis, 2002; Wilson and Szostak, 1999). *In vitro* selection has been used to identify aptamers to targets covering a wide range of sizes and functions, including proteins, antibiotics, sugars, vitamins, cofactors, nucleotides, nucleic acids, and amino acids (Hesselberth et

al., 2000; Wilson and Szostak, 1999; Famulok, 1999), simple ions (Ciesiolka and Yarus, 1996), small biomolecules (Ellington and Szostak, 1992), peptides (Nieuwlandt et al., 1995), proteins (Tuerk and Gold, 1990), organelles (Ringquist et al., 1995), viruses (Pan et al., 1995), and even entire cells (Morris et al. 1998) (An overview of aptamers generated by *in vitro* selection can be found in: Gold et al. (1995; Table 1)).

The first RNA aptamer targeted to a small biomolecule was directed at ATP (Sassanfar and Szostak, 1993), one of the most important metabolic cofactors.

The large number of small-molecule aptamers that have been isolated over the past several years permits some generalizations concerning the selection process and the resulting aptamers. First, RNA is clearly capable of recognizing a large variety of small molecules. Successful selections have been performed on a diverse array of small targets (Wilson and Szostak, 1999; Table 1), including both planar and nonplanar compounds having overall negative or positive charges. Surprisingly, even molecules that are largely hydrophobic, such as valine or tryptophan, can be recognized. Second, many of these RNA aptamers show impressive specificity. The theophylline aptamer for example discriminates against caffeine, a molecule possessing one additional methyl group, by at least 10,000-fold (Jenison et al., 1994).

A variation on the aptamer selection strategies can also be used to isolate catalytic nucleic acid sequences. Such approaches have been used to change the function of known ribozymes, to create new ribozymes, and to understand their structures, catalytic mechanisms, and even folding pathways (Michel et al., 1990; Green et al., 1990; Green and Szostak, 1992; Beaudry and Joyce, 1992; Bartel and Szostak, 1993b; Lehman and Joyce, 1993; Tsang and Joyce, 1994; Chapman and Szostak, 1995; Tsang and Joyce 1996; Hager and Szostak, 1997; Tuschl et al., 1998; reviewed in Jäschke et al., 1999).

To date, no DNA enzymes of natural origin have been found. However, an increasing number of catalytic DNAs, with characteristics that are similar to those of ribozymes, are being produced outside the confines of a cell (Schubert et al., 2003; reviewed in Breaker, 1997a).

Because proteins carry out a wider range of structural and catalytic roles in

biology and are much more extensively used in diagnostic, therapeutic and industrial applications, great interest has also been generated in the development of methods for *in vitro* selection and directed evolution of proteins (Roberts and Szostak, 1997; Keefe and Szostak, 2001; Lamla and Erdmann, 2003).

Recently, there also has been an increased interest in the application of aptamers as biosensors for the detection and measurement of biological or environmental ligands (Famulok and Mayer, 1999; Jhaveri et al., 2000; Osborne et al., 1997), and as *in vivo* probes of biological function (Famulok et al., 2000). In addition, aptamers that are active *in vivo*, so-called "riboswitches", have been discovered: These allosteric ribozymes are part of mRNAs and able to influence gene expression (Winkler et al., 2002).

Aptamers isolated by *in vitro* selection have not only broadened our insight into the capabilities of nucleic acids, but the isolation of diverse new ribozymes by *in vitro* selection also supports the notion that ribozymes could have directed a primitive metabolism before the evolution of protein synthesis (see reviews by Joyce, 2002 and Wilson and Szostak, 1999). By studying the interaction of aptamers with their targets, valuable insights can therefore be obtained into the structure-function relationship of natural and artificial biomolecules.