In *International Journal of Unconventional Computing*, Vol. 5(2), pp. 165-189, 2009, Old City Publishing, Inc.

Nucleic Acid Enzymes: The Fusion of Self-assembly and Conformational Computing

Effirul I. Ramlan and Klaus-Peter Zauner

School of Electronics and Computer Science
University of Southampton, SO17 1BJ, United Kingdom
{eir05r,kpz}@ecs.soton.ac.uk

Abstract. Macromolecules are the predominant physical substrate supporting information processing in organisms. Two key characteristics—conformational dynamics and self-assembly properties—render macromolecules unique in this context. Both characteristics have been investigated for technical applications. In nature's information processors self-assembly and conformational switching commonly appear in combination and are typically realised with proteins. At the current state of biotechnology the best candidates for implementing artifical molecular information processing systems that utilise the combination self-assembly and conformational switching are functional nucleic acids. The increasingly realised prevalence of oligonucleotides in intracellular control points towards potential applications. The present paper reviews approaches to integrating the self-assembly and the conformational paradigm with allosterically controlled nucleic acid enzymes. It also introduces a new computational workflow to design functional nucleic acids for information processing.

1 Biomolecular Computing Paradigms

With the feature size of solid-state devices approaching nanometer scale molecules are coming increasingly into focus as an alternative material substrate for the implementation of information processing devices.

A prominent difference between solid-state materials and macromolecular materials is the large range of properties found in molecules. Biomolecules are mainly composed from only six (C, H, O, N, S, P) out of the 91 naturally occurring chemical elements. The number of possible compounds that could in principle be formed from these six atoms is very large. Even though there are many restrictions on how the atoms can be combined, stable macromolecules comprising hundreds or thousands of atoms can be formed. Macromolecules occurring in organisms are typically formed from a set of building block molecules. These building blocks link through covalent bonds originating at specific atoms, but can be combined in arbitrary order. The twenty commonly occurring amino acids form such a set of building blocks. Linear polymers from up to a few

hundred of these amino acids linked in arbitrary sequential order, constitute an important class of biomacromolecules, the proteins. Another set of building blocks found in nature are the nucleotides which combine, again in arbitrary order, to long nucleic acid molecules. The exact linear sequence of the building blocks may have a relatively small influence on the properties of the complete macromolecule, as is the case with the deoxyribonucleic acids (DNA), the carriers of genetic information in the cell. But the exact sequence can also be crucial to the properties of the macromolecule, as is typical for proteins. Both cases have practical advantages. The former is ideally suited for representing information, because the physical properties of the macromolecule are largely independent of its information content. The latter case gives rise to the diverse specificity and large range of material properties that is the basis of the tremendous variety of organisms seen in nature.

Two phenomena are key to the interaction and function of macromolecules: self-assembly and conformational dynamics. Both play also an important role for molecular information processing in nature and each serves as a paradigm for man-made molecular computing schemes. Figure 1 illustrates these paradigms. Atoms attached through covalent bonds in a molecule can exert weak, short-

Fig. 1. Cartoon of the two basic biomolecular computing paradigms. In *self-assembly computing* (A) information is encoded by molecular shapes. Molecules with complementary shape form supra-molecular clusters through non-covalent binding. Thus shape-encoded input is mapped into features of the cluster as output. In *conformational computing* (B) the physicochemical environment of a macromolecule serves as input signal. Intramolecular dynamics maps this milieu information into a change in conformational state.

range attractive forces (van-der-Waals interaction, hydrogen-bonds) on atoms in other molecules. If two macromolecules have complementary surfaces, i.e., surfaces that allow for close proximity of a large number of suitable atom-pairs, then the additive effect of the weak attractive forces results in a stable binding of the complementary molecules. In other words, the potential energy will dominate entropy even at high temperature. Molecules, on the one side, are large enough to have specific shape features. On the other side, they are small enough to be moved around by thermal motion and therefore can explore each others shapes by diffusion.

The self-assembly paradigm (Fig. 1A) effectively converts a symbolic pattern recognition problem into a free-energy minimisation process [18–20]. The self-assembly paradigm can conveniently be implemented with deoxyribonucleic

acid (DNA) because it is relatively easy to predict the binding among oligonucleotides from thermodynamic data and computer simulations. Over the past decade a number of experimental realisations of self-assembly computing have been reported (e.g., [1, 46, 74]). A drawback of self-assembly computing is that the random search of molecules for complementary partners by means of Brownian motion does not scale well to large reaction volumes.

In conformational computing (Fig. 1B) one attempts to exploit the shape changes that large macromolecules can undergo in response to their environment. The freedom of atoms in a molecule to rotate around single covalent bonds equips molecules with considerable flexibility. In proteins in particular have a distinctive agility that is core to their folding from a linear amino acid chain to a compact functional or structural component. This flexibility, however, does not terminate with the folding. The physicochemical milieu in which the macromolecule is embedded modulates the transition probabilities among the molecule's conformational states. Different conformational states commonly result in altered functional activity. A few experimental implementations that make use of the conformational dynamics have been reported (e.g., [32, 76, 5]). In the conformational paradigm much of the computation is an intramolecular process and state changes can therfore be fast. However, a problem with this approach is that in practice the conformational effects are at least hard and often impossible to predict. The practical implementations so far rely on molecules occurring in nature or genetically engineered variants of these molecules [32].

In nature's molecular information processing infrastructure both self-assembly and conformational dynamics play an important role. Typically both occur in combination. A protein may undergo a conformational change and as a consequence of this its shape becomes complementary to a region on another macromolecule thus leading to self-assembly. Conversely, a molecule that participates in self-assembly experiences a significant change in its environment as a result of the binding to another molecule and this change can give rise to an altered conformation. In combination self-assembly and conformational switching are a powerful set of primitives on which the entire molecular machinery of cells is built. It would be desirable to combine the self-assembly and the conformational paradigm also for artificial molecular computing schemes. In nature proteins are the key components that integrate self-assembly and conformational switching. Unfortunately, both phenomena are notoriously difficult to predict for proteins. Recently an intriguing alternative has been experimentally demonstrated in form of nucleic acid enzymes, i.e., DNA or RNA molecules with catalytic activity. [62, 53].

The main goal of the present paper is to provide a self-contained overview of nucleic acid enzymes from an information processing perspective. The remainder of this paper will first introduce nucleic acid enzymes in general, next describe their application in information processing experiments, and finally turn to software tools supporting the design of nucleic acid enzymes. Moreover, a new work-flow for designing functional RNA is introduced.

2 Properties of Nucleic Acids

Nucleic acids are macromolecules that play an important role as information carriers in cells. Two types occur, ribonucleic acids (RNA) and deoxyribonucleic acids (DNA), which are named after the structure of a sugar component always present in these molecules. RNA and DNA are typically long linear polymers that consist of a large number of monomers taken from a set of four different nucleotides. The sequential order in which the nucleotides are interlinked in the nucleic acid molecule can represent information.

The two types of nucleic acids play different roles. RNA has the task of transmitting information within the cell, while DNA transmits information from generation to generation. The genetic information is encoded in a dimer of two complementary nucleotide chains ('single-stranded' DNA), which upon self-assembly assumes the well known double-helical structure ('double-stranded' DNA'). DNA is well suited as carrier of genetic information, because of its energy degeneracy with respect to the sequential order of the nucleotides. The properties of RNA molecules are more dependent on the sequence of nucleotides and as a consequence, RNA takes on additional roles in the cell aside from representing information.

Each of the monomer units that make up nucleic acids consists of a sugar moiety, a phosphate group, and a base. The sugar component of the monomers in RNA molecules is ribose, hence the name *ribonucleic acid*. Correspondingly, DNA is named *deoxyribonucleic acid* after its sugar component deoxyribose. The flexibility of macromolecules to change their three-dimensional shape, i.e. their *conformation*, while maintaining the covalent bonds among atoms, i.e. their *configuration* unchanged, is the basis of conformational computing. The chemical stability of DNA and the structural flexibility of RNA are both desirable properties for molecular computing based on nucleic acid enzymes. The type of nucleic acid that is preferred for the implementation of a particular molecular component, will often depend on this tradeoff.

Within either RNA or DNA, the sugar moieties and the phosphate groups of all monomers are identical. The base that forms the third component of each monomer provides the variety requisite for representing information in the sequence of monomers. Each monomer unit carries one of four possible bases. In RNA these are adenine, guanine, cytosine, and uracil, abbreviated as A, G, C, and U. The first three of these bases also occur in DNA, but instead of uracil DNA contains thymine (T). This difference is thought to be of use for DNA repair mechanisms that actively maintain the integrity of a cell's genetic information, but is of no relevance within the context of the present paper.

Of crucial importance for the interaction of nucleic acid molecules is the *complementarity of bases*. The base of a nucleotide can form weak bonds, called hydrogen-bonds, with another nucleotide that carries a complementary base. Hydrogen bonds occur between a hydrogen atom bound to an electronegative atom, and another electronegative atom. They are roughly 20× weaker than a covalent bond.

Among the four possible bases that can occur in a nucleotide, T or U can bind to A with two hydrogen bonds and G can bind to C forming three such bonds. Two nucleotide strands with complementary base sequences will form a dimer that is held together by the additive effect of the hydrogen bonds that can be formed between the complementary bases. This process is called *hybridisation*. The direction of the sequence has to be taken into account if complementarity is considered. The two strands that form a double helix are intertwined running in opposite direction. To indicate the orientation of a single stranded nucleic acid, its ends are named after the unbound carbon atom in the sugar moiety as $5'$ at one end and $3'$ at the other. As a convention, the notation of nucleic acid sequences is written from left to right in $5'$ to $3'$ direction (i.e., ATTGC always stands for $5'$–ATTGC–$3'$) [7]. In the following figures a diamond symbol ($\diamond$) indicates the $3'$-end of a strand.

If a nucleic acid has a sequence that is complementary to itself, then it can fold back onto itself and form an intramolecular double-helix [55]. Partial intramolecular hybridisation can result in a complex three-dimensional structure of the molecule. In some instances the three-dimensional structure confers functionality such as a specific catalytic activity.

As mentioned above, RNA is more flexible than DNA and as a consequence it forms more readily intramolecular base-pairs. A single stranded RNA molecule can bind to itself in several regions, with the unbound segments present as loops between bound segments or dangling ends. The loops can be grouped into four classes, illustrated in Fig. 2. Due to its higher flexibility, in addition to the pair-

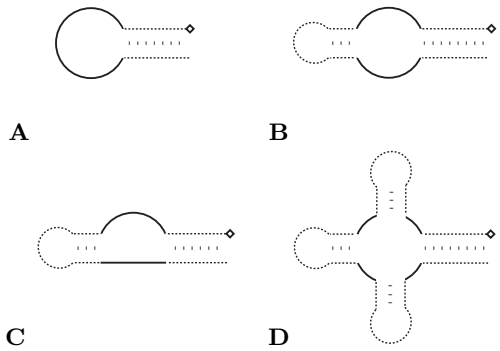


Fig. 2. Classification of RNA loop motifs; the named motif is shown with solid lines. Hairpin loop (A), internal loop (B), bulge (C), multi-branch loop (D; a four-way junction is shown). A hairpin loop with the adjacent stem is referred together as stem loop. After [70].

ing of complementary bases (A-U/U-A, C-G/G-C), the 'wobble pairing' of G-U (and reverse oriented U-G) through two hydrogen-bonds also contributes to the structural variability exhibited by RNA molecules [71]. For a given RNA sequence ('primary structure'), there is often a diverse set of secondary structures

it can fold into. Which structure is favoured will depend on the environment of the molecule, for example, the presence or absence of other molecules or ions. Conversely, a diverse set of sequence configurations can yield a particular secondary structure [23, 77]. Subsequently, interactions among secondary structure motifs lead to the formation of a tertiary structure, which in some cases entails functionality. Such functional RNA molecules are discussed in the next section. A review of RNA secondary structure and its prediction is offered by [34].

3 Functional Nucleic Acids

Biological catalysis was thought to be synonymous with catalytically active protein, i.e., enzymes, until RNA molecules with catalytic capability were discovered [2]. These *ribozymes*, as the RNA enzymes are also called, led to the hypothesis that precursors of the cell may have relied on RNA only for the both transmission of genetic information and metabolism, tasks which are in present cells relegated to DNA and protein, respectively. Although it appears unlikely that DNA has catalytic function in nature, it was possible to arrive at DNA enzymes in the laboratory [10].

Ribozymes can be categorised according to size and catalytic activity [22, 65]. The three classes of ribozyme are small catalytic RNAs, group I and II introns, and Ribonuclease P (RNase P). Small catalytic RNAs range in size from 40–160 nt (nucleotides) and are self-cleaving molecules. The group of small catalytic RNA comprises of hammerhead ribozymes [9], hairpin ribozymes [22], the hepatitis delta virus (HDV) ribozyme [28], and the Neurospora Varkud Satellite ribozyme [42]. Each of these ribozymes has a distinct structure. Nevertheless, all of them catalyse the same reaction. They cleave the phosphodiester bond in RNA, generating a $5'$-product with a $2'$, $3'$-cyclic phosphate terminus and a $3'$-product with a $5'$-hydroxyl terminus. It is thought that the $2'$ hydroxyl group of the ribose moiety of RNA participates in the catalysis [27], however DNA can also act as a catalyst, as will be discussed later in this section.

Most of the known natural occurring ribozymes catalyse intramolecular (also called *in-cis*) reactions, in which the ribozyme cuts and detaches from part of its own sequence [30]. However, some ribozymes have been successfully modified to split other nucleic acids. To avoid ambiguity, we will use the term *ribozyme core* to refer to the catalytically active RNA molecule in intermolecular (*in-trans*) reactions. Such a reaction is illustrated in Fig. 3. The figure shows the sequence of reaction steps in the catalytic cycle of a hammerhead ribozyme. For brevity, the release of both products is shown as a single step, however, the products are likely to dissociate from the ribozyme core one after another [45]. The turnover rate of small ribozymes is typically about 1 cleavage per minute [73, 22].

For some of the ribozyme cores it is feasible to control their catalytic activity. This property is key to the application of ribozymes in molecular computing. In order to understand the mechanisms of controlling the activity of the ribozymes, it is useful to consider their secondary structure. The secondary structure that emerges from an RNA sequence is composed from the motifs in Fig. 2 and

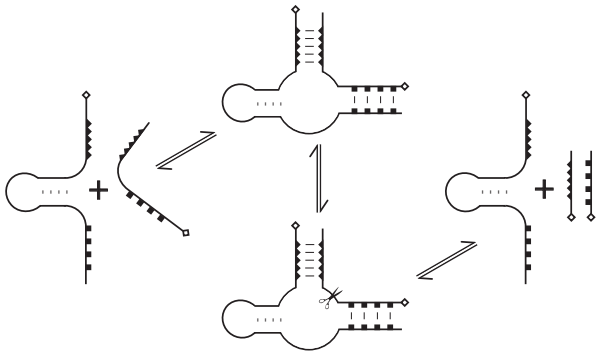


Fig. 3. Splitting of an RNA molecule catalyzed by another RNA molecule [45]. The catalytic RNA binds a substrate RNA molecule if it is complementary to the two hybridisation regions indicated by squares and triangles (left). The substrate molecule is cut at the location indicated by scissors (centre). After the two reaction products dissociate from the catalytic RNA, the latter is ready for another reaction cycle (right). The 3′-ends of the RNA sequences are indicated by diamonds (⋄).

possibly dangling single-stranded ends. The different types of ribozymes are distinguished by their characteristic combination of loops and helices [44]. The secondary structure of a hammerhead ribozyme is depicted in Fig. 4. Hammerhead ribozymes require the presence of a metal ion (typically Mg^{2+}) to be catalytically active [22]. For convenience the description of these structures is given in the figure captions. A more detailed discussion of their mechanisms can be found in [21, 65, 69].

It is generally believed that the conformational flexibility of RNA is important for the catalytic process itself [21, 36]. The conformational flexibility RNA also gives rise to a large variety of secondary structures. The secondary structure consists of single stranded regions alternating with double stranded regions where stretches of the RNA molecule bind to itself (cf. Figs. 4). These motifs interact and from the 3-dimensional tertiary structure of the RNA molecule. The conformational flexibility thus supports a diverse set of functional roles [50].

The structural variety of RNA and its concomitant functional diversity make RNA a suitable medium for directed in-vitro evolution [37]. This technique is based on the possibility to copy RNA molecules with aid of protein enzymes. Errors in the copy process yield a population of RNA molecules with slightly varied sequences. Repetitive application of this error-prone replication process, will lead to an evolution of the population of molecules. In the absence of other selection pressures, the evolution would favour molecules that are most efficiently reproduced by the participating protein enzymes. However, a selection step can be introduced to assert evolutionary pressure in another direction. The molecules could for example be selected by their binding capabilities towards a particular substrate molecule [26].

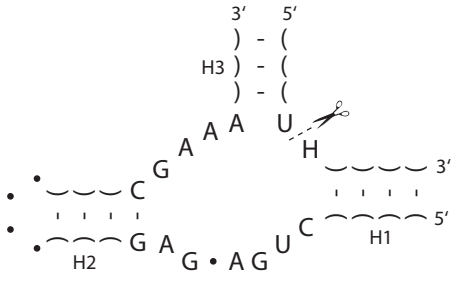


Fig. 4. Minimal functional structure of hammerhead ribozyme. Three helical stems (H1, H2, H3) emanate from a junction on the ribozyme core [33, 66]. In nature, always either helix H1 or H3 is terminated by a hairpin loop which results in intramolecular catalysis. Hammerhead ribozymes that catalyse the *in-trans* reaction, as depicted in the figure, can be made synthetically [9]. The core region has a specific sequence for all known active structures and is therefore termed 'conserved'. Conserved bases are specified explicitly, with H representing any one of {A, C, U}. A dot (•) stands for any base that will not cause hybridisation in this position; correspondingly two parentheses connected by a dash indicate an arbitrary pair of complementary bases.

A number of ribozymes have been produced through directed evolution [13, 17]. The majority of them possess a ribozymes core that does not resemble any of those found in nature [68]. Directed evolution provides a technique to enrich the repertoire of RNA structures amenable to molecular computing applications.

Directed evolution can also be applied to DNA and, rather surprisingly, yields DNA molecules with enzymatic activity (deoxyribozymes) [12, 14, 56, 38]. DNA is best known as a memory molecule inscribed with information crucial for the production of macromolecular components in cells. The properties that make DNA suitable for this function are its stability, reliable hybridisation, but also the fact that DNA forms a double-helical structure largely independent of the sequence of bases as long as the two strands that hybridise are complementary. These properties together with the absence of DNA enzymes in nature, had led to the view that DNA is not flexible enough to act as a catalyst. It is now, however, well established that DNA does have the structural flexibility to support a range of secondary and tertiary structures [58] and can form a diverse set of tertiary structures with a potential to function as catalysts [10]. Secondary structures of three deoxyribozymes developed through the process of in-vitro selection are depicted in Fig. 5. For the deoxyribozymes shown in panels B and C of Fig. 5, it was found that their catalytic reaction rates are comparable to those of ribozymes [25]. As mentioned above, hammerhead ribozymes require the presence of metal ions to be catalytically active. The deoxyribozymes also require metal ions. The first deoxyribozyme was designed in the presence of Pb^{2+} as cofactor [12] and the deoxyribozymes shown in Fig. 5 all require Mg^{2+}.

From an application perspective, the use of DNA has the advantage over RNA that DNA molecules are generally more stable. Furthermore, the DNA-

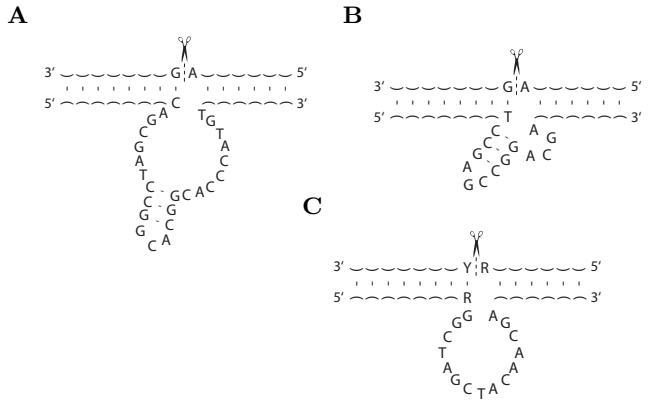


Fig. 5. Secondary structures of three deoxyribozymes. Panel A shows a deoxyribozyme that resembles the general structure of the hammerhead ribozymes (cf. Fig. 4). It is characterised by a specific ('conserved') region of 15 nt connected to a stem-loop and is capable of cleaving substrates that contain a G-A-joint. A deoxyribozyme with a different structure, but also applicable only to substrates sequences with a G-A-joint is shown in panel B. The deoxyribozymes in panels A and B both have been applied for information processing [62]. Panel C shows a deoxyribozyme that is less constrained in the substrate junction it will cleave [56]. The notation for the binding region is: Y∈{C, T} and R∈{A, G}.

DNA-binding is more reliable and results in higher specificity. Given these practical advantages of DNA and the fact that DNA enzymes do not occur in nature, it is of particular interest that recently a ribozyme was successfully converted into a deoxyribozyme by means of directed evolution [51]. A DNA sequence that corresponded (apart from the T for U substitution) to a known ribozyme which catalyses a covalent bonding between two RNA oligonucleotides was found to be inactive. However, after acquiring suitable mutations during directed evolution, a deoxyribozyme that also catalyses a covalent bonding between two RNA oligonucleotides—though at a lower efficency and different bond location—was arrived at.

In combination, the capability of self-assembly through hybridisation of complementary sequences and the conformational flexibility to form sequence dependent spatial structures with catalytic activity, make nucleic acids an attractive material for molecular computing.

4 Nucleic Acid Enzymes for Computing

From the time it became apparent that nucleic acid polymers carry the genetic information in their base sequence, the astounding information density was recognised. Early suggestions for implementing a molecular computer with DNA followed the encoding principle of genetic information (cf. [43]). This would require the formation and cleavage of numerous covalent bonds for their operation

and thus require specific sets of enzymes. Major progress in the application of nucleotides for information processing came about two decades later with Adleman's insight that random oligonucleotides could be the basic tokens for information processing [1]. His method employed enzymes only to stabilise (through covalent bonds) the products of a self-assembly process (hybridisation of partially complementary oligonucleotides), but not in the information processing itself, and accordingly did not require enzymes with sequence specificity.

More recently these two lines of thought have come together with the use of nucleic acid enzymes [62–64, 53]. Key to this approach is the possibility to control the activity of a ribozyme or deoxyribozyme with oligonucleotides. Such allosterically controlled nucleic acid enzymes have been investigated as sequence specific biosensors, where they have the advantage over molecular beacons that they catalytically amplify the recognition event [41]. The concept of allosteric ribozymes is illustrated in Fig. 6. Within certain constraints, the base sequence for the binding site of the control oligonucleotide (labeled OBS in the figures) can be chosen of the sequence on which the nucleic acid enzyme will act. It is therefore possible to have an oligonucleotide sequence start (or stop) the production of another, largely independent, oligonucleotide sequence. Moreover, it is possible to engineer nucleic acid enzymes to be controlled by more than one oligonucleotide. For instance, the molecule shown in Fig. 7 was designed by adding an allosteric control to the deoxyribozyme shown in Fig. 5B [62]. It is inactive unless two effector molecules with specific base sequences are present. The behaviour of the molecule can be interpreted as an AND logic gate. Note, however, that the possibility to catalyse the production of oligonucleotides as output signal with a

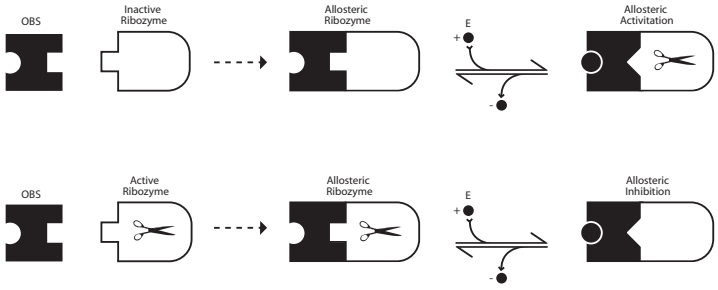


Fig. 6. Allosterically activated ribozyme (top) and allosterically inhibited ribozyme (bottom) [61, 59]. The allosteric ribozyme is composed of two components (left of the dashed arrow), a oligonucleotide binding site (OBS) and a ribozyme part. The two components are covalently bound and from a single nucleic acid molecule (centre). Upon binding an effector oligonucleotide (E) the conformation of the binding site changes and affects the conformation of the ribozyme component. The latter conformational change will activate (top) or inhibit (bottom) the catalytic activity of the ribozyme part. The same scheme can also be realised with deoxyribozymes.

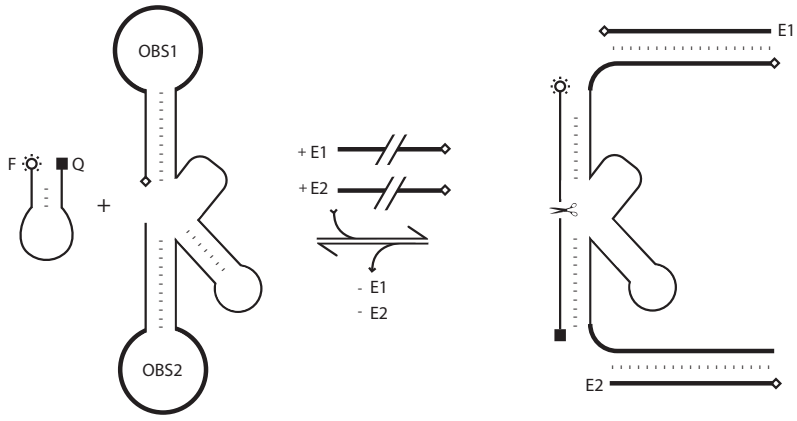


Fig. 7. Deoxyribozyme acting as a logic AND gate after [62]. The molecule is designed in such a way that it can self-hybridise to block its own substrate binding site. This self-hybridisation is weaker than the binding of the effector molecules (E1, E2) to their oligonucleotide binding sites (OBS1, OBS2). Only in the presence of both effector molecules is the substrate binding site accessible and accordingly the deoxyribozyme catalytically active. By supplying a molecular beacon (far left) as substrate the output of the gate can be determined optically. If the deoxyribozyme is catalytically active it will cleave the beacon molecule, thus separate the quencher (Q) from the fluorophore (F), and consequently give rise to a fluorescence signal.

base sequence independent of the sequences that serve as input signals (effector molecules) allows for applications other than logic AND operations.

A hammerhead ribozyme requiring the presence of two specific oligonucleotides for it to become active is shown in Fig. 8. While the deoxyribozyme gate in Fig. 7 is inactivated by blocking the substrate binding site, the ribozyme in Fig. 8 is controlled by a different mechanism. In the absence of effector oligonucleotides, the molecule will self-hybridise to form a structure that is not a ribozyme. Hybridisation with the effector molecules overcomes the self-hybridisation of the inactive conformation and the molecule changes into a structure with a hammerhead ribozyme component. A comparison of the the multi-branch loop on the far right of Fig. 8 with the structural requirements of a hammerhead ribozyme depicted in Fig. 4 reveals how the straitening of the oligonucleotide binding sites upon hybridisation with two DNA effector molecules induces catalytic activity.

The conformational dynamics of RNA molecules allows for a relatively straight forward design of allosteric control structures into known ribozymes along the line of the concept represented in Fig. 6. Accordingly, hammerhead ribozymes have been engineered with a wide variety of effector molecules [61]. One strategy is to add effector binding sites at the crucially important helix II of the hammerhead structure (cf. Fig. 4). Due to the conformational flexibility of RNA, it

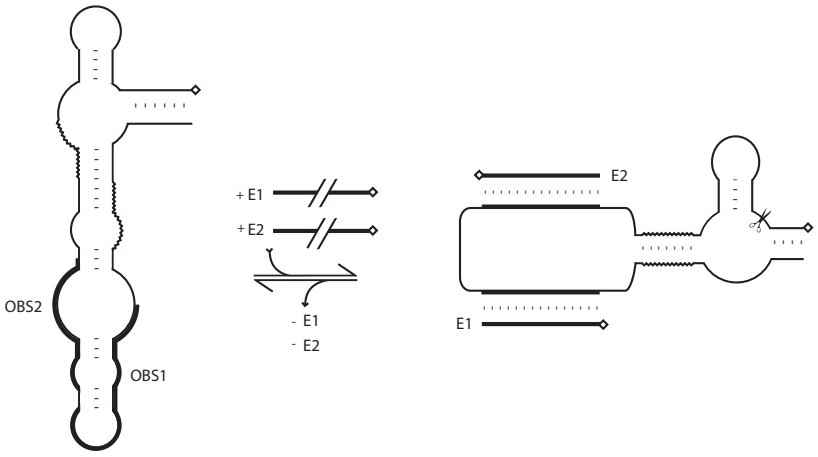


Fig. 8. Two-input molecular switch based on allosterically controlled hammerhead ribozyme after [53]. In the absence of effector molecules (E1, E2) the inactive conformation (left) is more stable. Upon binding of the effector molecules to their corresponding oligonucleotide binding sites (OBS1, OBS2) the ribozyme changes into a catalytically active conformation (right). Crucial for the formation of the hammerhead conformation is the correct self-hybridisation in the helix II region shown by crinkled lines in both conformations. The oligonucleotide binding sites are indicated by bold lines in both conformations.

is then likely that an effector molecule binding to the ribozyme will affect the helix II conformation and thus disrupt the catalytic function.

For the application of ribozymes as signal processing components, RNA structures that can be controlled with nucleic acid oligonucleotides as effector molecules are of particular interest. This is the case because the control oligonucleotide may conceivably be the product of a reaction catalyzed by another ribozyme and therefore enable the implementation of small molecular control networks. Different approaches to controlling a hammerhead ribozyme by means of oligonucleotide effectors are illustrated in Fig. 9. All four have been demonstrated in experiments [54, 15, 40, 72]. The first three (A–C) follow a common design philosophy. Starting from the basic hammerhead ribozyme structure shown in Fig. 4, an RNA sequence is engineered that does not fulfil the requirements for a hammerhead ribozyme, but can overcome this deficiency by hybridising with an effector oligonucleotide.

The earliest implementation of an engineered allosteric control mechanism in a ribozyme [54] is based on an RNA molecule that can form a hammerhead ribozyme, but has a preferred secondary structure that does not resemble the hammerhead motif and shows no catalytic activity; Fig. 4A. The self-hybridisation that stabilises the preferred conformation (left side in Fig. 4A.) can be overcome by a suitable effector molecule, the binding of which is energetically more favourable than the self-hybridisation. Upon binding the effector molecule, the

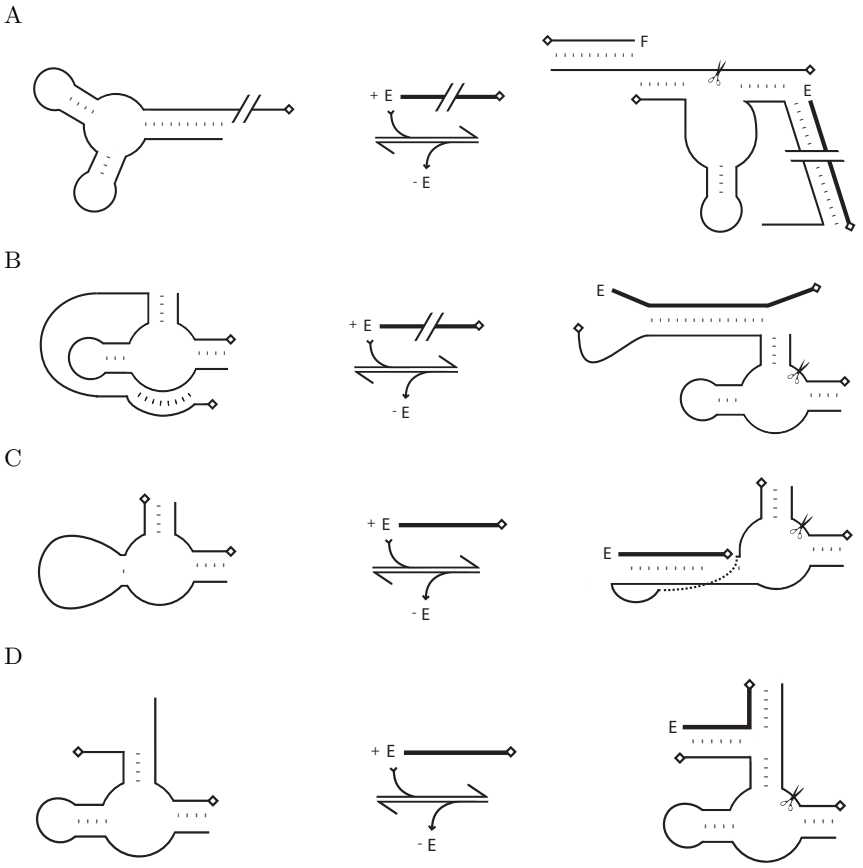


Fig. 9. Four different strategies to control a hammerhead ribozyme. In all cases the ribozyme is active only in the presence of an oligonucleotide effector (E). Panel A: Formation of hammerhead structure upon binding of effector [54]; a DNA facilitator strand (F) enhances the binding of substrate to ribozyme [31]. Panel B: Effector releases conserved core junction from hybridisation [15]. Panel C: Effector enables formation of helix II [40]. Panel D: Effector supports binding of substrate [72].

RNA sequence folds into an active hammerhead conformation. This control strategy is the one that has been used in the molecular switch shown in Fig. 8 [53].

The hammerhead motif of the ribozyme in Figure 9B is inactivated by self-hybridisation between the 3′-end of the ribozyme and its conserved junction region [15]. Between the region of the ribozyme participating in helix III and the region near the 3′-end that is complementary to part of the conserved core is an effector binding site. The binding of an oligonucleotide effector to the binding region is energetically favoured over the self-hybridisation in the core region. Accordingly, the binding of the effector releases the hybridisation of the core and activates the hammerhead structure. As mentioned earlier, the helix II is

a necessary part of the hammerhead motif and its stability is important for the enzymatic activity of hammerhead ribozymes [9]. Fig. 4C shows a control strategy based on an RNA sequence that contains the essential components of a hammerhead motif short of the complementary regions that could form the helix II. Binding of the effector induces a pseudo-half-knot structure that together with the helix formed between the effector strand and the ribozyme apparently forms a pseudo-stem capable of activating the ribozyme [40].

In contrast to the three allosteric ribozymes just described, the ribozyme shown in Fig. 4D is always in a catalytically active state and can cleave a sequence that will bind fully to form helix I and helix III (cf. Fig. 4 for helix positions). However, the catalytic activity with regard to substrate sequences that bind only partially in the helix III region can be controlled by an effector molecule [72]. The effector binds to the dangling $3'$-end of the ribozyme and the dangling $5'$-end of a suitable substrate. It thus facilitates the binding between the ribozyme and a substrate that would not be cleaved without the effector. Note, that the effector sequence in this case will influence on which substrates the ribozyme acts.

The combination of molecular motifs found in nature, molecules developed through directed evolution, and rational design decisions have led to a set of allosterically controlled functional nucleic acids suitable as components for simple molecular information processors. From the diverse family of catalytically active nucleic acids that have been found [26, 24, 68], it appears likely that the set of available components will grow.

5 Design of Functional RNA for Computation

Nucleic acids cannot compete with proteins when it comes to conformational dynamics or self-assembly . But from the practical perspective of implementation, using nucleic acids has the important advantage that both their conformations and their self-assembly can be predicted far more readily than the behaviour of proteins.

Computational tools are available for predicting the likely secondary structure of RNA molecules and for suggesting sequences that can be expected to fold to a desired structure. In particular, the prediction of RNA secondary structure has received much attention in the literature and mature software is available for this task. *MFOLD* [78] and *RNAfold* in the *Vienna Package* [35] are commonly applied. They differ in the model used to calculate the free energy of folded RNA structures, with the former using nearest neighbour rules to assign free energies to loops, while the latter employs a partition function [48] to estimate the probabilities of folding alternatives. Both programs can also be employed for predicting DNA conformation by substituting the appropriate thermodynamic parameters [55].

With a view to computing, predicting the folding of a single nucleotide strand in isolation, however, is not sufficient. As exemplified by the allosteric regulation mechanisms discussed in the previous section, the secondary structure which will

be assumed by a nucleic acid strand may depend on the presence or absence of other oligonucleotides. Predicting the interaction among multiple nucleic acid molecules is therefore of particular interest. The *MFOLD* [47] program and the *Vienna Package* [49, 8] offer the prediction of RNA-RNA and DNA-DNA co-folding, but lack the thermodynamic parameters for predicting RNA-DNA interactions. The *RNAsoft* package [3] allows for predicting the interaction of multiple RNA strands.

For designing RNA molecules with specific properties, it is often useful if a secondary structure can be specified and a suitable nucleotide sequence that will fold into the specified structure is generated from the specification. This so called inverse folding problem, is addressed by *RNAinverse* [35], *RNAdesigner* [3], and *INFORNA* [16]. Generally there is a large number of possible sequences that will fold into a given secondary structure; the aforementioned programs typically provide one arbitrary sequence from this pool.

By combining these tools Penchovsky and Breaker devised a workflow to engineer information processing units from nucleic acid sequences [53]. They used this workflow to design the biomolecular AND gate illustrated in figure 6. In the following we will describe the approach of [53] in more detail and subsequently discuss a modification to the workflow that places less restrictions on the inactive conformation. The design of a simple gate will serve to exemplify the approach and to investigate the design space. The simplest logic gates have only one bit input. Two such gates exit. The NOT gate that inverts the input and the PASS gate (sometimes also called "identity" or YES gate) that forwards the input signal. From a purely logical viewpoint, PASS gates serve no purpose, in practice they can reform a degraded signal or adjust signal delay [75]. The molecular pass gates considered below are more powerful than one-bit logic gates and an essentially arbitrary input sequence of limited length can be recoded into a different output sequence.

The workflow employed by Penchovsky and Breaker to engineer an RNA PASS gate [53] starts with a design for an approximately 80 nucleotides long RNA molecule which contains a highly sensitive hammerhead ribozyme in its sequence. The bases for the sequence of the molecule are fixed, except for a region of about 10–20 nucleotides long for which is crucial for maintaining the active hammerhead conformation. If this region can participate in internal hybridisation, the molecule will undergo a large change in conformation to a catalytically inactive state. The binding of an effector oligonucleotide to this region will prevent internal hybridisation and thus stabilise the catalytically active conformation of the hammerhead ribozyme. Accordingly, the region acts as an oligonucleotide binding site (OBS, cf. section 4) that exerts allosteric control over the catalytic activity of the ribozyme.

The aim is to design the sequence for this OBS, such that it is likely to allow the switching between the active and inactive state in a real RNA molecule. This is achieved by first selecting a candidate sequence for the OBS and inserting it into the fixed sequence of the sensitive hammerhead ribozyme. The sequence is generated by randomly assigning bases to the positions in the sequence while

Table 1. The constraints imposed on potential candidate sequences following [53].

Stage Filter	*Condition to satisfy*
1 Identical nucleotides	No more than three identical consecutive nucleotides in the oligonucleotide binding site(s)
2 Active state conformation	Active hammerhead conformation in the presence of effector
3 Base-pairing percentage	In the absence of effector(s) 30%–70% of the oligonucleotide binding region is hybridised
4 Energy gap	Energy gap between the inactive and active state is within -6 kcal/mol to -10 kcal/mol
5 Temperature tolerance	Structure is preserved over a temperature range of 20°–40°C
6 Ensemble diversity	For neither active nor inactive state the ensemble diversity (cf. [52]) exceeds 9 units
7 Folding efficiency	The RNA molecule must fold, in the absence of the effector, to the inactive conformation within 480 units in *Kinfold* [29]

obeying the constraint in the first row of Tab. 1. To judge the plausibility for this RNA sequence design to be practicable and likely to be operative if implemented as a real RNA molecule Penchovsky and Breaker introduced a filter cascade, the steps of which are summarised as rows two to six in Tab. 1. If a generated sequence passes these five filter steps, it is taken as a model design for the secondary structure of the OBS region in the desired gate. This model is specified by the complete secondary structure and a partial sequence, which commits to all bases except those located in the the OBS region. By repeatedly running *RNAinverse* with this specification one obtains a set of complete sequences for the logic gate which differ from each other only in the OBS region. Note that the secondary structure of sequences generated by *RNAinverse* may not strictly conform to the specified conformation, but does not differ by more than two base pairs. Only sequences that have a thermodynamic stability comparable to the model design are maintained. The folding efficiency of these sequences is then verified (last row of Tab. 1). In [53] a second stage of processing is suggested. It is applied to the sequence designs that have successfully passed the filter cascade and yields alternative sequences with similar folding and thermodynamic stability is suggested. In our tests the yield of this second stage was only about 3% better than that of the first stage. The value of the second stage presumably lies in providing sequence alternatives to designs that are already favoured,

e.g., because they have been verified experimentally. In contrast to the highly constraint design procedure of Penchovsky and Breaker [53] outlined above, we introduce now a protocol for designing RNA gates. This protocol relaxes the constraints on the secondary structure of the inactive conformation a gate can assume. Yet, in general, the length of functional nucleotide sequences composed of four bases in arbitrary order gives rise to a combinatorially large design space in which a random search without appropriate constraints would not be efficient in generating useful designs. However, the design space spanned by the RNA gate designs found in the literature can be used to narrow the search. In Tab. 2, the properties of existing implementations of RNA logic gates are summarised. The parameter ranges derived from the nucleic acid gates described in the literature are shown in Tab. 3; these values were used in the design of the PASS gates described below.

Our protocol for designing an allosterically controlled ribozyme comprises three generating steps, outlined in Tab. 4, to arrive at a sequence design. The generation of the sequence is followed by a series of validation steps. In the first step the conformation for the catalytically active ribozyme is determined by specifying the secondary structure of an extension to the hammerhead core composed of helix II, two linkers and an OBS region. To this end for each position in the sequence its participation in internal hybridisation is selected by generating a dot-bracket representation for the secondary structure of the molecule. In generating the secondary structure, constraints derived from Tab. 2 and detailed in Tab. 3, are invoked by a generation algorithm that follows [4].

Table 2. Structural elements of biomolecular logic gates designed from deoxyribozymes and ribozymes. For OBS, helices, hairpin loops, bulges, and internal loops the entries in the table give the size (in nucleotides) of the structural element.

Gate	*Length*	*OBS*	*Number of Junctions*	*Helices*	*Hairpin loops*	*Bulges*	*Internal loops*	Ref.
$PASS_1$	60	15	2	3, 7	3, 15	-	-	[62]
$PASS_2$	63	15	2	5, 6	4, 15	-	-	[64]
$PASS_3$	80	22	3	5, 8, 16	4, 7	1, 1	3	[53]
$PASS_3$	80	22	3	5, 8, 16	4, 7	1	4	[53]
NOT_1	52	15	1	5	15	-	-	[62]
NOT_2	92	22	3	4, 10, 13	6, 6	-	2, 4, 8	[53]
AND_1	78	15, 15	3	3, 8, 9	3, 15, 15	5	-	[62]
AND_2	85	15, 15	3	5, 6, 6	4, 15, 15	-	-	[64]
AND_3	73	15, 15	2	8, 9	15, 15	-	-	[63]
AND_4	112	16, 16	3	5, 8, 21	4, 7	1	2, 5, 10	[53]
OR	103	20, 20	3	5, 8, 23	4, 7	1, 1	2, 4, 6	[53]
a AND $\neg b$	76	15,15	2	5,7	15,15	-	-	[62]
a AND b AND $\neg c$	96	15,15,15	3	5,6,6	15,15,15	-	-	[64]

Table 3. Parameters for structure generation derived from Tab. 2.

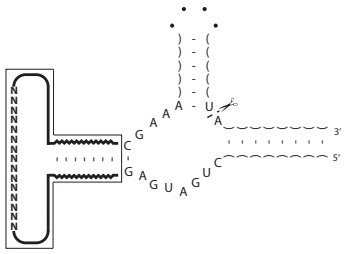
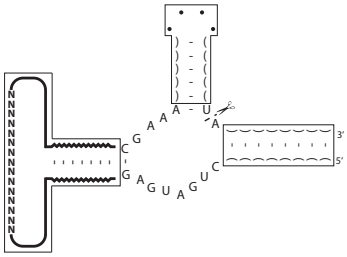
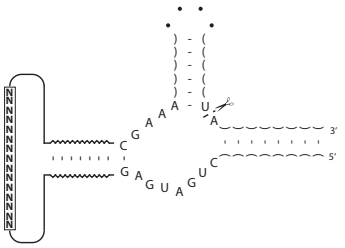
Type	Probability	Maximum no.	Length Range
Helix	0.50	-	4–15
Hairpin Loop	-	0–3	4–15
Internal Loop	0.45	0–3	2–8
Bulge	0.05	0–1	1–8
Junction	-	0–3	4–8
OBS	-	1	15–22
Linker	0.55	2	0–5

The second step of the protocol assigns nucleotides to the positions in the sequence, except the OBS region. This assignment of the nucleotides adheres to the secondary structure generated in the first step. It repeatedly uses a simplified version of the initialisation routine of *RNAdesigner*. Rather than applying the subsequent refinement of the nucleotide assignment offered by *RNAdesigner* [4], we reject sequence assignments that do not fold to the specified secondary structure in *RNAfold* and repeat the initialisation. For the folding tests of the nucleotide assignment, the unassigned OBS region is set to a repetition of a hypothetical non-binding base as suggested by Penchovsky and Breaker [53].

The hammerhead ribozyme can reliably be deactivated by binding to its conserved core-region and thus distorting its secondary structure [67, 39, 61, 11]. Therefore, in the third generating step of then protocol, first a sequence complementary to the conserved CUGAUGAG-region of the hammerhead core is inserted at a random location within the two linkers and the OBS, i.e., in the hairpin loop attached to helix II. Afterwards, the remaining unassigned positions in the sequence are filled by drawing randomly from the four possible nucleotides {A, U, C, G}. The resulting sequence is likely to be inactive due to internal hybridisation of a section of the OBS, and possibly a few bases from a linker, to the hammerhead core. In the folding predictions, a hybrid molecule composed of ribozyme and substrate is considered and an interference of the substrate with the OBS is unlikely.

The subsequent validation of the generated sequence designs involves all steps of the filter cascade in Tab. 1. This screening process prunes out 99% of the generated sequence candidates as depicted in Fig. 10. Starting with a pool of 58 423 candidates as input to the first filter stage, 586 designs passed the entire filter chain. A manual inspection of the dot-plot graph [35] for all 586 designs confirmed in every case that the conserved sequence region of the hammerhead core is blocked by hybridisation in the inactive conformation and free in the active conformation. To further evaluate the plausibility of the remaining computational designs, we calculate the equilibrium constants for the three possible dimers that can form when ribozyme molecules and effector molecules interact. The calculation is based on the free-energy values provided by *RNAcofold* [35, 49] and the assumption of a fixed, equal concentration for the monomeric ribozyme (R) and monomeric effector (E) [8]. Any point in the area of the triangle

Table 4. Proposed computational procedure for designing allosterically controlled hammerhead ribozyme gates. In contrast to the method described in [53] the procedure starts with the conformation of the active ribozyme.

	Design of the secondary structure of a potential gate based on a hammerhead core. First the structure (but not sequence) of the helix II and associated allosteric control domain is generated. This extension of the ribozyme is comprised of four parts: a helix that attaches to the ribozyme core, an effector binding region (NN$\cdots$NN), and two linker sequences connecting binding site and helix. The lengths of the extension (23–62 nt) and the helix (4–15 nt) are chosen randomly. The remaining part of the structure (binding region and linkers) is filled with the constraints listed in Tab. 3.
	Assign the conserved nucleotides of the hammerhead core. The remaining sequence positions except the binding site are assigned by searching for a base sequence that will fold into the structure designed in the previous step. This can be achieved with *RNAinverse from Vienna* [35], *RNAdesigner from RNAsoft* [3, 4], or *INFORNA* [16]. To arrive at the active conformation of the gate, a non-binding pseudo-base (N) is assigned to all positions in the binding region during the search process.
	Replace the pseudo-bases in the binding region with real bases. This is done such that the structural elements initially generated for the linker and binding parts of the extension become manifest. This can be verified with *MFOLD* [78] or *RNAfold from Vienna* [35].

depicted in Fig. 11 corresponds to a calculated (cf. Eq. 13 of [57]) combination of the relative concentrations of the three possible dimers that can form (RR,

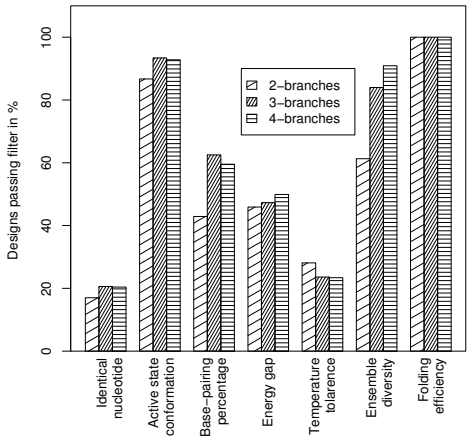


Fig. 10. Permissiveness of each step in the filter chain of Tab. 1 when applied to generated candidate sequences for PASS gates. The filter stages are applied consecutively from lefty to right. For each step the sequences adhering to the filter condition is shown as percentage of input sequences supplied to this filter stage.

EE, RE). From this analysis, it appears that the enlarged degrees of freedom in the design procedure outlined in Tab. 4, can yield PASS-gate designs with good ribozyme-effector binding (RE). Note however, that due to the lack of tools for simulating RNA-DNA hybridisation, the values shown for the designs from Penchovsky and Breaker have been calculated for an RNA effector, while in [53] an experimentally more convenient DNA effector molecule was applied.

Samples of structures that were derived with the procedure outlined in table 4 and have passed the screening with the filter chain in table 1, are shown in Fig. 12. The three structures in panel A, B, and C are representative for the classes of molecules that have inactive conformations with 2-branches, 3-branches, and 4-branches, respectively. In each structure the oligonucleotide binding region for the effector molecule is indicated by a bold line section. The structure in panel A bears some resemblance to a design proposed by Porta and Lizardi [54] (cf. Fig. 9A), while the structure in panel B is similar to the designs by Penchovsky and Breaker [53].

6 Conclusion

Ribonucleic acids are an attractive computing substrate for three reasons. Firstly, they support both basic paradigms of molecular computing: self-assembly and conformational switching. Secondly, the intramolecular and intermolecular hybridisation of nucleic acids can reasonably well be predicted with existing computational tools. Thirdly, nucleic acids play a very important role for memory and control in every living cell. Ribonucleic acids are also challenging to work with in the laboratory, however. Yet, these challenges are at present more manageable than the computational challenge one would face if one would attempt

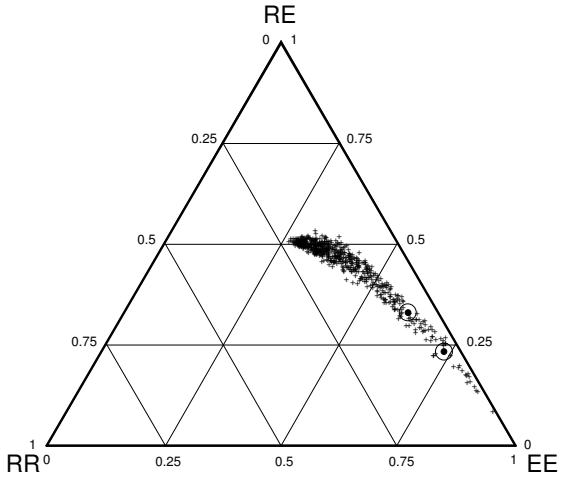


Fig. 11. Estimated binding between ribozyme (R) and effector (E) for different PASS gate designs. The plot indicates the values obtained for PASS gates that have been designed with the method of table 4 and evaluated with the filter chain in table 1 labeled as +, and the two experimentally validated designs from [53] labeled ⊙.

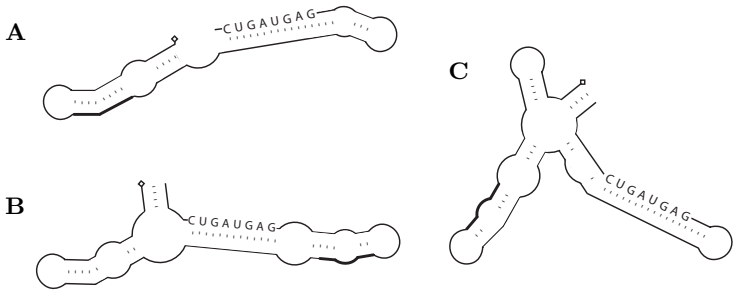


Fig. 12. Inactive conformation of allosterically controlled hammerhead ribozymes designed to act as PASS gates.

to design information processing components with nature's main information processing substrate, namely proteins.

The application of allosterically controlled nucleic acids in bioimmersive computation has the potential to open up extraordinary possibilities. Smart drugs that can sense the internal state of cell and intervene in the intracellular regulatory mechanisms may come within reach [6] and engineered molecular control mechanisms that can be integrated into cells would be a powerful tool for life-science research [60].

Before the potential of these long-term aims can be realised, obstacles in the laboratory need to be tackled and much better computational design procedures

are required. A crucial issue will be the prediction of the interactions within complex mixtures of molecules. At present secondary structure prediction for multiple RNA strands is at its infancy and simulation tools capable of predicting DNA-RNA interactions do not exist. It will be necessary to develop a general methodology and supporting computational tools to create purpose-designed sets of interacting allosterically controlled nucleic acids.

Acknowledgements

E.R. is grateful to the Ministry of Higher Education Malaysia and the University Malaya for financial support.

References

1. L. M. Adleman. (1994). Molecular computation of solutions to combinatorial problems. *Science*, 266:1021–1024.
2. S. Altman. (1990). Enzymatic cleavage of RNA by RNA (Nobel lecture). *Angewandte Chemie International Edition*, 29:749–758.
3. M. Andronescu, R. Anguirre-Hernández, A. Condon, and H. H. Hoos. (2003). RNAsoft: A suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Research*, 31(13):3461–3422.
4. M. Andronescu, A. P. Fejes, F. Hutter, A. Condon, and H. H. Hoos. (2004). A new algorithm for RNA secondary structure design. *Journal of Molecular Biology*, 336(3):607–624.
5. R. Baron, O. Lioubashevski, E. Katz, T. Niazov, and I. Willner. (2006). Elementary arithmetic operations by enzymes : A model for metabolic pathway based computing. *Angew. Chem. Int. Ed.*, 45:1572–1576.
6. Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro. (2004). An autonomous molecular computer for logical control of gene expression. *Nature*, 429:423–429.
7. J. M. Berg, J. L. Tymoczko, and L. Stryer. (2003). *Biochemistry*. W. H. Freeman and Company, New York, fifth edition.
8. S. H. Bernhart, H. Tafer, U. Mückstein, C. Flamm, P. F. Stadler, and I. L. Hofacker. (2006). Partition function and base pairing probabilities of RNA heterodimers. *Algorithms for Molecular Biology*, 1(3). doi: 10.1186/1748-7188-1-3.
9. K. R. Birikh, P. A. Heaton, and F. Eckstein. (1997). The structure, function and application of the hammerhead ribozyme. *Eur. J. Biochemistry*, 245:1–16.
10. R. R. Breaker. (1997). DNA enzymes. *Nature Biotechnology*, 15:427–431.
11. R. R. Breaker. (2002). Engineered allosteric ribozymes as biosensor components. *Current Opinion in Biotechnology*, 13:31–39.
12. R. R. Breaker and G. F. Joyce. (1994). A DNA enzyme that cleaves RNA. *Chemistry & Biology*, 1:223–229.
13. R. R. Breaker and G. F. Joyce. (1994). Inventing and improving ribozyme function: Rational design versus iterative selection methods. *Trends in Biotechnology*, 12:268–275.
14. R. R. Breaker and G. F. Joyce. (1995). A DNA enzyme with Mg^{2+}-dependent RNA phosphoesterase activity. *Chemistry & Biology*, 2:655–660.
15. D. H. Burke, N. D. S. Ozerova, and M. Nilsen-Hamilton. (2002). Allosteric hammerhead ribozyme TRAPs. *Biochemistry*, 41:6588–6594.

16. A. Busch and R. Backofen. (2006). INFO-RNA—a fast approach to inverse RNA folding. *Bioinformatics*, 22(15):1823–1831.
17. K. B. Chapman and J. W. Szostak. (1994). In vitro selection of catalytic RNAs. *Current Opinion in Structural Biology*, 4:618–622.
18. M. Conrad. (1989). Self-assembly as a mechanism of molecular computing. In *Proceedings of the 11th Annual International IEEE-EMBS Conference*, pages 1354–1355, Piscataway, NJ. IEEE.
19. M. Conrad. (1992). Molecular computing: The lock-key paradigm. *Computer (IEEE)*, 25(11):11–20.
20. M. Conrad. (1993). Emergent computation through self-assembly. *Nanobiology*, 2:5–30.
21. E. A. Doherty and J. A. Doudna. (2000). Ribozyme structures and mechanisms. *Annual Review of Biochemistry*, 69:597–615.
22. J. A. Doudna and T. R. Cech. (2002). The chemical repertoire of natural ribozymes. *Nature*, 418:222–228.
23. D. E. Draper. (1996). Strategies for RNA folding. *Trends in Biochemistry Sciences*, 21:145–149.
24. A. D. Ellington and J. W. Szostak. (1990). In vitro selection of RNA molecule that bind specific ligands. *Nature*, 346:818–822.
25. G. M. Emilsson and R. R. Breaker. (2002). Deoxyribozymes: New activities and new applications. *Cellular and Molecular Life Sciences*, 59:596–607.
26. M. Famulok and J. W. Szostak. (1992). In vitro selection of specific ligand-binding nucleic acids. *Angewandte Chemie International Edition*, 31:979–988.
27. M. J. Fedor and J. R. Williamson. (2005). The catalytic diversity of RNAs. *Molecular Cell Biology*, 6:399–412.
28. A. R. Ferré-D'Amaré, K. Zhou, and J. A. Doudna. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature*, 395:567–574.
29. C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster. (2000). RNA folding at elementary step resolution. *RNA*, 6:325–338.
30. A. C. Forster and R. H. Symons. (1987). Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell*, 49:211–220.
31. J. Goodchild. (1992). Enhancement of ribozyme catalytic activity by a contiguous oligodeoxynucleotide (facilitator) and by 2'-O-methylation. *Nucleic Acids Research*, 20(17):4607–4612.
32. N. Hampp. (2000). Bacteriorhodopsin as a photochromic retinal protein for optical memories. *Chem. Rev.*, 100:1755–1776.
33. K. J. Hertel, D. Herschlag, and O. C. Uhlenbeck. (1994). A kinetic and thermodynamic framework for the hammerhead ribozyme reaction. *Biochemistry*, 33:3374–3385.
34. P. G. Higgs. (2000). RNA secondary structure: physical and computational aspects. *Quarterly Reviews of Biophysics*, 33(3):199–253.
35. I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. (1994). Fast folding and comparison of RNA secondary structures. *Chemical Monthly*, 125(2):167–188.
36. S. Hohng, T. J. Wilson, E. Tan, R. M. Clegg, D. M. J. Lilley, and T. Ha. (2004). Conformational flexibility of four-way junctions in RNA. *J. Mol. Biol.*, 336:69–79.
37. G. F. Joyce. (1992). Directed molecular evolution. *Scientific American*, 267(6):48–55.
38. G. F. Joyce. (2004). Directed evolution of nucleic acid enzymes. *Annual Review of Biochemistry*, 73:791–836.

39. M. Koizumi, G. A. Soukup, J. N. Q. Kerr, and R. R. Breaker. (1999). Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nature Structural Biology*, 6(11):1062–1071.
40. Y. Komatsu, S. Yamashita, N. Kazama, K. Nobuoka, and E. Ohtsuka. (2000). Construction of new ribozymes requiring short regulator oligonucleotides as a cofactor. *Journal of Molecular Biology*, 229:1231–1243.
41. T. Kuwabara, M. Warashina, and K. Taira. (2000). Allosterically controllable ribozymes with biosensor functions. *Current Opinion in Chemical Biology*, 4:669–677.
42. D. A. Lafontaine, D. G. Norman, and D. M. J. Lilley. (2001). Structure folding and activity of the VS ribozyme: Importance of the 2-3-6 helical junction. *EMBO Journal*, 20(6):1415–1424.
43. E. A. Liberman. (1979). Analog-digital molecular cell computer. *BioSystems*, 11:111–124.
44. D. M. J. Lilley. (1999). Structure, folding and catalysis of the small nucleolytic ribozymes. *Current Opinion in Structural Biology*, 9:330–338.
45. D. M. Long and O. C. Uhlenbeck. (1993). Self-cleaving catalytic RNA. *FASEB Journal*, 7:25–30.
46. C. Mao, T. H. LaBean, J. H. Reif, and N. C. Seeman. (2000). Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature*, 407:493–496.
47. N. R. Markham. (2003). Hybrid: a software system for nucleic acid folding, hybridizing and melting prediction. Master's thesis, Rensselaer Polytechnic Institute.
48. J. S. McCaskill. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119.
49. U. Mückstein, H. Tafer, J. Hackermüller, S. Bernhard, P. F. Stadler, and I. L. Hofacker. (2006). Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22(10):1177–1182. doi:10.1093/bioinformatics/btl024.
50. K. Nagai and L. W. Mattaj. (1994). *RNA-Protein Interactions*. Oxford University Press, New York, first edition.
51. N. Paul, G. Springsteen, and F. Joyce. (2006). Conversion of a ribozyme to a deoxyribozyme through in vitro evolution. *Chemistry & Biology*, 13(3):329–338.
52. R. Penchovsky and J. Ackermann. (2003). DNA library design for molecular computation. *Journal of Computational Biology*, 10(2):215–229.
53. R. Penchovsky and R. R. Breaker. (2005). Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nature Biotechnology*, 23(11):1424–1433.
54. H. Porta and P. M. Lizardi. (1995). An allosteric hammherhead ribozyme. *Bio/Technology*, 13:161–164.
55. J. SantaLucia, Jr. and D. Hicks. (2004). The thermodynamics of DNA structural motifs. *Annual Review of Biomolecular Structure*, 33:415–440.
56. S. W. Santoro and G. F. Joyce. (1997). A general purpose RNA-cleaving DNA enzyme. In *Proceedings of the National Academy of Sciences, USA*, volume 94, pages 4262–4266.
57. P. Schuster. (2006). Prediction of RNA seconday structures: from theory to models and real molecules. *Reports on Progress in Physics*, 69:1419–1477.
58. N. C. Seeman. (2003). DNA in a material world. *Nature*, pages 427–431.
59. S. K. Silverman. (2003). Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA*, 9:377–383.
60. M. L. Simpson. (2004). Rewiring the cell: synthetic biology moves towards higher functional complexity. *Trends in Biotechnology*, 22(11):555–557.

61. G. A. Soukup and R. R. Breaker. (1999). Nucleic acid molecular switches. *Trends in Biotechnology*, 17:469–476.
62. M. N. Stojanovic, T. E. Mitchell, and D. Stefanovic. (2002). Deoxyribozyme-based logic gates. *Journal of the American Chemical Society*, 124:3555–3561.
63. M. N. Stojanovic and D. Stefanovic. (2003). Deoxyribozyme-based half-adder. *Journal of the American Chemical Society*, 125:6673–6676.
64. M. N. Stojanovic and D. Stefanovic. (2003). A deoxyribozyme-based molecular automaton. *Nature Biotechnology*, 21(9):1069–1074.
65. R. H. Symons. (1992). Small catalytic RNAs. *Annual Review of Biochemistry*, 61:641–671.
66. R. H. Symons. (1997). Plant pathogenic RNAs and RNA catalysis. *Nucleic Acids Research*, 25(14):2683–2689.
67. J. Tang and R. R. Breaker. (1997). Rational design of allosteric ribozymes. *Chemistry & Biology*, 4:453–459.
68. J. Tang and R. R. Breaker. (2000). Structural diversity of self-cleaving ribozymes. In *Proceedings of the National Academy of Sciences, USA*, volume 97, pages 5784–5789.
69. N. K. Tanner. (1999). Ribozymes: The characteristics and properties of catalytic RNAs. *FEMS Microbiology Reviews*, 23:257–275.
70. I. Tinoco, Jr. and C. Bustamante. (1999). How RNA folds. *Journal of Molecular Biology*, 293:271–281.
71. G. Varani and W. H. McClain. (2000). The G·U wobble base pair. *EMBO reports*, 1(1):18–23.
72. D. Y. Wang, H. Y. Lai, A. R. Feldman, and D. Sen. (2002). A general approach for the use of oligonucleotide effectors to regulate the catalysis of RNA-cleaving ribozymes and DNAzymes. *Nucleic Acids Research*, 30(8):1735–1742.
73. J. E. Wedekind and D. B. McKay. (1999). Crystal structure of lead-dependent ribozyme revealing metal binding sites relvant to catalysis. *Nature Structural Biology*, 6(3):261–268.
74. E. Winfree. (2000). Algorithmic self-assembly of DNA: Theoretical motivations and 2d assembly experiments. *J. of Biomolecular Structure & Dynamics*, 11(2):263–270.
75. K.-P. Zauner. (2005). Molecular information technology. *Critical Reviews in Solid State and Material Sciences*, 30(1):33–69.
76. K.-P. Zauner and M. Conrad. (2001). Enzymatic computing. *Biotechnology Progress*, 17(3):553–559.
77. M. Zuker. (1989). On finding all suboptimal foldings of an RNA molecule. *Science*, 244:48–52.
78. M. Zuker, D. H. Mathews, and D. H. Turner. (1999). *RNA Biochemistry and Biotechnology*, chapter Algorithm and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide, pages 11–43. NATO ASI. Kluwer Academic Publishers, Dordrecht.