A small catalytic RNA motif with Diels–Alderase activity

Burckhard Seelig and Andres Jäschke

Background: The 'RNA world' hypothesis requires that RNA be able to catalyze a wide variety of chemical reactions. *In vitro* selection from combinatorial RNA libraries has been used to identify several catalytic activities, most of which have resulted in a self-modification of RNA at one of its constituents. The formation of carbon–carbon bonds is considered an essential prerequisite for a complex metabolism based on RNA.

Results: We describe the selection and characterization of new ribozymes that catalyze carbon–carbon bond formation by Diels–Alder reaction of a biotinylated maleimide with an RNA-tethered anthracene. Secondary structure analysis identified a 49-nucleotide RNA motif that accelerates the reaction about 20,000-fold. The motif has only 11 conserved nucleotides that are present in most of the selected sequences. The ribozyme motif is remarkably adaptable with respect to cofactor and metal-ion requirements. The motif was also re-engineered to give a 38-mer RNA that can act as a 'true' catalyst on short external substrate oligonucleotide–anthracene conjugates.

Conclusions: We have identified a small, highly abundant RNA motif that can solve the complex task of forming two carbon–carbon bonds between two reactants *in trans*, a catalytic capacity useful for creating prebiotically relevant molecules. This is the smallest and fastest RNA catalyst for carbon–carbon bond formation reported to date.

Address: Institut für Biochemie der FU Berlin, Thielallee 63, 14195 Berlin, Germany.

Correspondence: Andres Jäschke
E-mail: jaschke@chemie.fu-berlin.de

Key words: cycloaddition, *in vitro* selection, oligonucleotide conjugates, RNA catalysis, ribozymes

Received: **25 November 1998**
Revisions requested: **30 December 1998**
Revisions received: **7 January 1999**
Accepted: **8 January 1999**

Published: **19 February 1999**

Chemistry & Biology March 1999, **6**:167–176
http://biomednet.com/elecref/1074552100600167

Introduction

Combinatorial RNA libraries have found increasing use for the identification of new ligands and catalysts [1,2]. The interest in isolating new ribozyme activities has grown tremendously, not only to explore the catalytic potential and the limits of RNA catalysis but also to test the hypothesis that, at some time in the past, life was based on RNA [3]. Consequently, there has been a big effort to expand the scope of RNA catalysis from reactions in which the substrates are nucleic acids to other basic chemical reactions. The direct selection method has been successfully used to isolate numerous self-modifying RNAs; some of these RNAs have been subsequently engineered to act as true catalysts (for recent reviews, see [4,5]).

The Diels–Alder cycloaddition is one of the central reactions in organic synthesis and, despite its minor biological significance, it has often been used in innovative approaches in catalysis research [6–8]. In the context of the 'RNA world' hypothesis, the formation of carbon–carbon bonds would be essential in anabolic pathways. Several laboratories have attempted, with little success, to generate RNA Diels–Alderase enzymes by selecting RNA aptamers against the respective transition-state analogs (e.g., [9]). Recently, a moderate Diels–Alder reaction catalyst consisting of RNA modified with additional functionalities attached to the nucleobases was reported [10]. To date, no catalysis of C–C bond formation by unmodified RNA has been described.

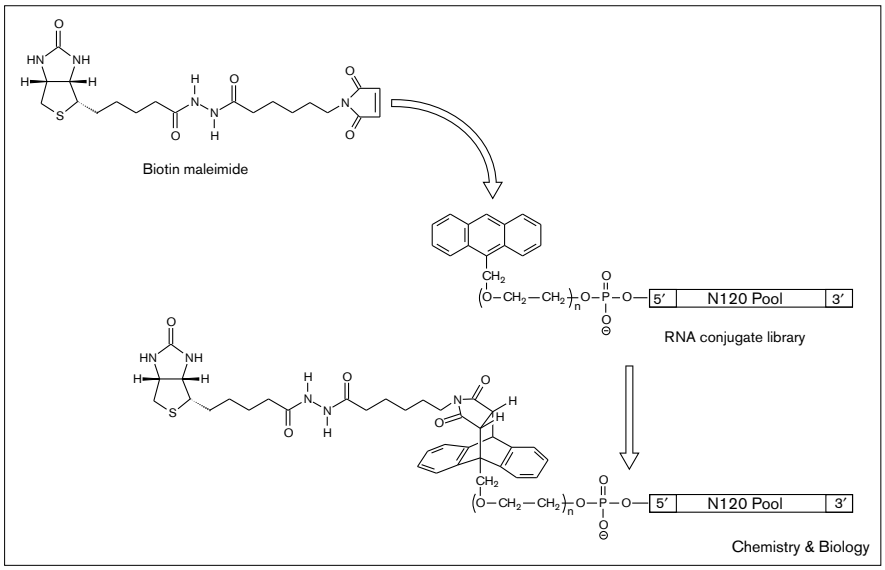
To expand our knowledge of RNA catalysis, we designed an *in vitro* selection scheme to isolate RNA molecules with Diels–Alderase activity (Figure 1). Our strategy involved generating a randomized pool of RNA conjugates in which anthracene as the diene is attached to the individual pool molecules via a long flexible polyethylene glycol (PEG) tether [11,12]. The resulting pool of anthracene–PEG–RNA conjugates was then incubated with biotinylated maleimide as the Diels–Alder dienophile to select RNA molecules that catalyze C–C bond formation between the two reactants. We anticipated that the reactive double-bond system of anthracene would be the main position to react with the maleimide. The newly formed biotinylated Diels–Alder product would then be covalently attached to the RNA through the PEG linker. After removal of all nonbiotinylated members of the pool using streptavidin agarose chromatography, active RNA molecules could be amplified and used in iterative selection cycles.

Results

Library construction

We generated a library of RNA conjugates with 120 randomized nucleotides, flanked by 20-nucleotide constant primer-binding sites. The starting library contained ~2×10^{14} sequence variants, and about 50 copies of each species were present. Anthracene–PEG was attached to the 5′ end of the RNA molecules by transcription initiation with ternary conjugates of guanosine monophosphate, PEG and anthracene

Figure 1



Reaction step in the protocol for the selection of Diels–Alderase ribozymes. Catalytic RNA molecules in the RNA–PEG–anthracene conjugate library catalyze the formation of two C–C bonds between the double-bond system of anthracene and (biotinylated) maleimide. Biotinylated product molecules are partitioned using streptavidin agarose, and the RNA part amplified enzymatically.

as described previously [11,12]. Polydisperse PEG with an average number of 13 ethylene glycol units was used. Typically, about 50% of the RNA transcripts were conjugated under the conditions used. After electrophoretic purification, the pools were used in the selection experiments.

Background reaction

The Diels–Alder reaction between anthracene and maleimide has been known to occur at a measurable rate in aqueous solution. Rideout and Breslow [13] reported a second-order rate constant of $0.226 \pm 0.007\ M^{-1}s^{-1}$ (H_2O, 45°C) for the reaction of 9-hydroxymethylanthracene with *N*-ethylmaleimide. To establish the rate of the uncatalyzed background reaction in our system, we prepared short constant-sequence anthracene–PEG–RNA conjugates, as well as unmodified oligonucleotides by T7 transcription, and incubated them in the presence of various metal ions with different concentrations of biotin maleimide. Electrophoretic separation of the reaction mixtures gave $k_{uncat} = (1.3 \pm 0.1) \times 10^{-2}\ M^{-1}s^{-1}$ under the selection conditions. No reaction of the unmodified oligonucleotides was observed under these conditions, supporting our assumption that the anthracene moiety would be the preferred reaction site.

Selection

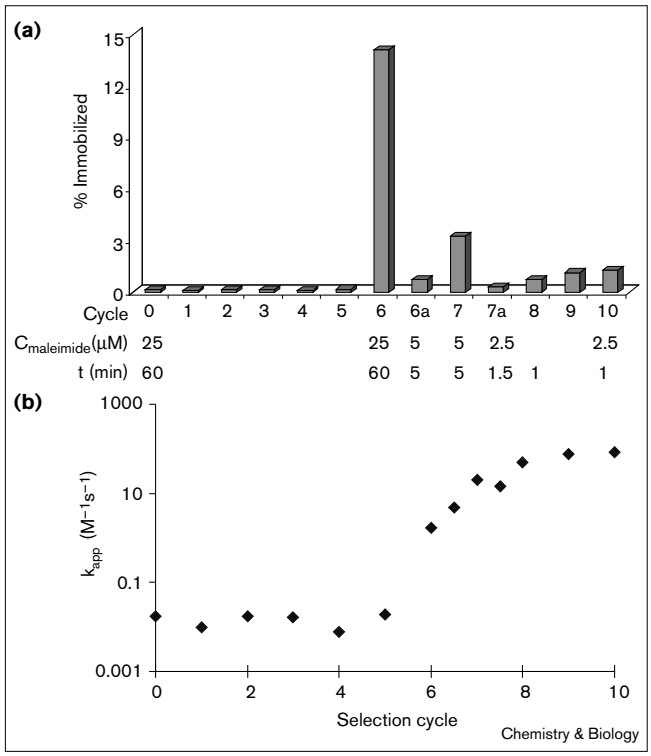
In the first six cycles of *in vitro* selection, 25 μM of biotin maleimide was incubated with 500 nM of anthracene–PEG–RNA. Several potential cofactors, like Lewis acids, metal ions, a dipeptide compound and a dipyridyl compound, were present in the reaction mixture. The incubation time was adjusted to 1 h, so that we could expect about 0.1% conversion due to the uncatalyzed background reaction. Consequently, in the first five rounds of selection we always selected between 0.1 and 0.2% of the library by immobilization on streptavidin agarose. The bound RNAs were reverse-transcribed and amplified using the polymerase chain reaction (PCR). *In vitro* transcription of the resulting DNA yielded an enriched pool of RNA conjugates that was used as input for the next round of selection. To prevent the enrichment of streptavidin binders, we carried out a preselection against streptavidin agarose prior to the addition of biotin maleimide in every other round of selection. Figure 2 shows the progress of the selection. Apparent rate constants (k_{app}) were calculated from the fraction of immobilized RNA in each selection round. In round 6, the fraction of immobilized RNA increased to 14%, indicating a 130-fold rate acceleration. Repetition of this round without anthracene–PEG–GMP initiator nucleotides gave less than 0.1% immobilized RNA, indicating that anthracene is the reaction site. Additionally, spectrometric monitoring of the disappearance of the typical anthracene fluorescence (419 nm) after addition of biotin maleimide demonstrated that the reaction occurs at the chromophoric system, most likely at the conjugated double-bond system in the central anthracene ring.

In rounds 6–10, the selection pressure was gradually increased; in round 10, the reaction time was only 1 min and the maleimide concentration was reduced to 2.5 μM. The k_{app} for the enriched pool in round 10 was $85\ M^{-1}s^{-1}$, which corresponds to a 6500-fold rate acceleration.

Cloning and sequence analysis

After round 10, the selected RNA was reverse-transcribed and PCR-amplified, and the resulting DNA was cloned. 55 clones were directly amplified using either the M13 forward and reverse primers to produce sequencing templates, or

Figure 2



(a) Enrichment of biotinylated RNA at each round of selection. Concentrations of biotin maleimide and reaction times are given below. In the first five rounds the amount of selected RNA was between 0.1 and 0.2% of the library. **(b)** Ribozyme activity increased with successive rounds of selection.

the specific primers used in the selection for the synthesis of individual RNA species by T7 transcription. A total of 42 different sequences were found, 35 of which gave at least a 3000-fold rate acceleration in the fluorescence spectrometric assay (Figure 3). 13 sequences accelerated the reaction by more than 15,000-fold. The best selected sequences had a k_{app} of ~240 $M^{-1}s^{-1}$, corresponding to a 18,500-fold rate acceleration.

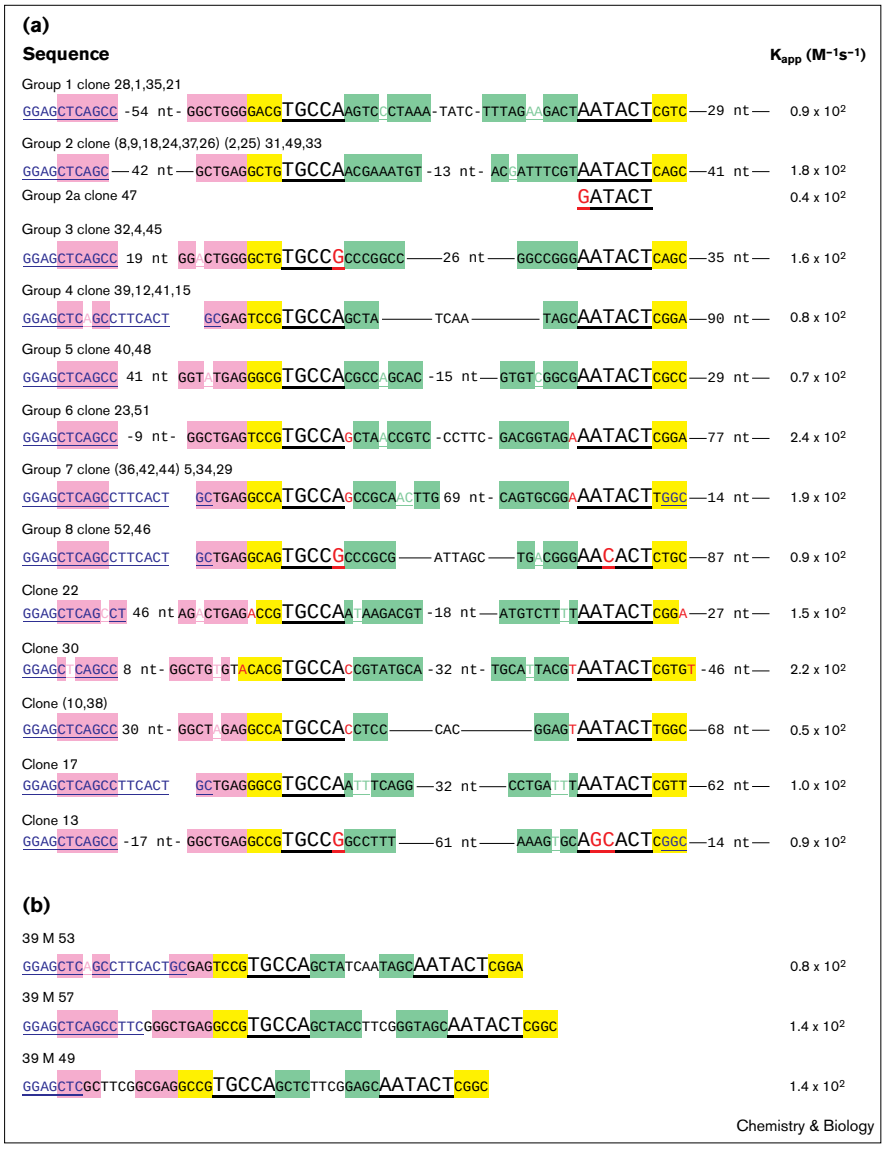
The most abundant individual sequence was found in six different clones. Primary sequences were highly diverse, and homologies were identified using the clustalw cluster alignment algorithm in the multalin implementation [14]. The active sequences could be assigned to eight families, and eight other sequences showed no homologies to each other or to the sequence families. Pair-wise comparison of the consensus sequences of the individual families identified a motif shared by ~90% of all active sequences: a pentanucleotide stretch with the sequence UGCCA in conjunction with a hexanucleotide stretch AAUACU. Several minor mutations of these sequences were also observed. For each sequence, the neighborhood of these stretches was pair-wise complementary, suggesting the formation of a double strand with an extended bulge in the middle. In addition, a nucleotide stretch complementary to the 5′ constant primer region was found directly upstream of these sequence elements.

These structure proposals were then confirmed by secondary-structure analysis using RNAdraw and RNA structure software [15,16]. The general motif shared by the majority of the sequences is shown in Figure 4, along with a proposed numbering scheme. The central element is the large bulge comprised of the two consensus sequences UGCCA (B1) and AAUACU (B2). Below the bulge, a double-stranded helix (helix II) is formed, which is always 4 base pairs (bp) long. The stem sequences are variable and change pair-wise, but preferences can be noted for each position. Above the bulge, another helix (helix III) is formed with variable length and with occasional mismatches and looped-out bases and bulges. In helix III, we did not detect any sequence preferences. In several clones, right above the bulge is a $G_{13.1}$–$A_{14.1}$ mismatch, leading to an enlarged bulge. In two sequences, we find a C–U mismatch at the same position. The upper ends of helix III are closed by a loop, L2, that is variable in length and sequence.

From helix II, a third helix (helix I) continues in which one strand is formed by at least five nucleotides from the conserved 5′-primer-binding site. The sequence of helix I is therefore rather conserved. Helices I and II are contiguous, and they are likely to be coaxially stacked. The ends of helix I are closed by a loop, L1, of variable size and sequence. In all sequences containing the motif, the 5′-terminal GGAG is left unpaired. The first G is the site where the linker-coupled anthracene is attached, so it is likely that this tetranucleotide, together with the linker, positions the anthracene so that it can reach inside the bulge, which might form the active site.

In several clones, secondary-structure analysis suggests a fourth helical region formed by folding back the 3′-terminal primer-binding region onto sequence stretches downstream of nucleotide 21.1. The sequence and length of this helix, its distance from helix II, and the sequence of the spacing region were highly diverse, however, indicating that these parts are not relevant for catalysis. Moreover, the combination of sequence information and apparent rate constants leads to the following conclusions. Firstly, mutations outside the motif shown in Figure 4 have no influence on the catalytic activity. Secondly, there are no simple relationships between the length, number of mismatches, or thermodynamic stability of helix I and III, and the catalytic rate acceleration. Thirdly, the five sequences with the G–A mismatch at the bottom of helix III (13.1–14.1) seem to be more efficient catalysts than those with the perfect helix. Fourthly, inside the bulge, nucleotides U_8, G_9, C_{10} and C_{11} in B1 seem to be absolutely conserved, whereas the replacement of A_{12} by G is tolerated; the replacement

Figure 3



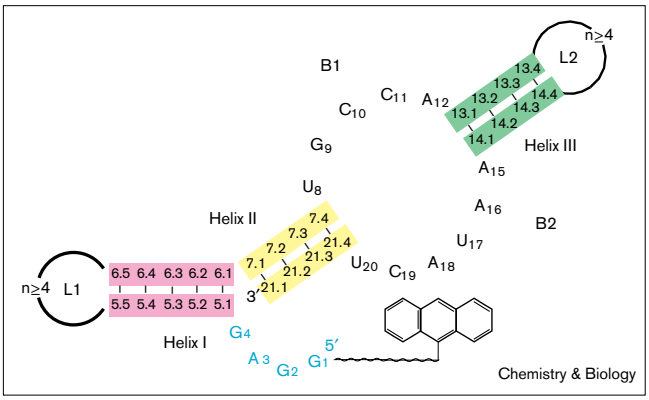
(a) Active sequences selected and apparent rate constants. **(b)** Minimized sequences. Constant primer sequences, blue and underlined; bulges, black, bold and underlined; helix 1, pink; helix 2, yellow; helix 3, green. Mismatches inside a helix are shown as interruptions of colored bars, and critical mutations are shown in red. All individual clones were assayed. Clone numbers in parentheses represent identical sequences. Assay conditions were 500 nM anthracene–RNA, 50 μM biotin maleimide, selection buffer, 2% DMSO, 23°C. Components were mixed and the decrease of fluorescence emission at 419 nm after excitation at 253 nm was monitored. Initial velocities were obtained graphically.

of A_{15} by G leads to a 75% decrease in activity (compare clone 47 with all other members of group 2); a U_{17}→C mutation and in one case an A_{18}→G mutation are tolerated; and A_{16}, C_{19} and U_{20} are absolutely conserved. Finally, in three sequences, we find an interesting double mutation: A_{12}→G is accompanied by U_{17}→C. In three other cases the A_{12}→G mutation occurs alone, but, interestingly, U_{17}→C does not occur alone. This might suggest a base-pairing interaction between these two positions.

Rational design and ribozyme engineering

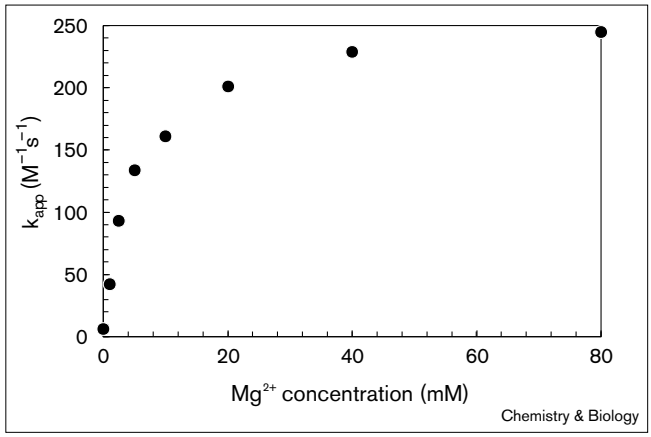
The information compiled in Figure 3 allowed us to proceed directly with condensed minimum structures rather than individual full-length transcripts. We prepared a 53-nucleotide truncated version of clone 39 RNA (39M53; Figure 3) using nested PCR and T7 transcription, because this medium-activity clone contains a condensed version of the active motif near the 5′ end. 39M53 showed the same catalytic activity (81 $M^{-1}s^{-1}$) as the nontruncated parental sequence 39, supporting our assumptions. Next, we rationally designed minimum structures with increased thermodynamic stability. We eliminated unpaired bases from the helices, and we replaced loops 1 and 2 with the stabilizing tetraloop sequences UUCG. For each position in helix II we chose the preferred nucleotide from Figure 3. The bottom of helix II was closed by a G7.1–C21.1 base pair, because this was expected to result in high stabilization when coaxially stacked on helix I [17]. Two different versions were prepared (Figure 3); in 39M57, the 57-nucleotide-long RNA has a 7 bp helix I and a 6 bp helix III,

Figure 4



Proposed catalytic motif and numbering scheme of the Diels–Alderase ribozyme motif. Helix I, pink; helix II, yellow; helix III, green; constant primer sequences, blue; B, bulge; L, loop.

Figure 5



The ribozyme activity of the engineered molecule is dependent on Mg^{2+}. Apparent rate constant values were obtained from fluorescence spectrometric measurements (see the Materials and methods section).

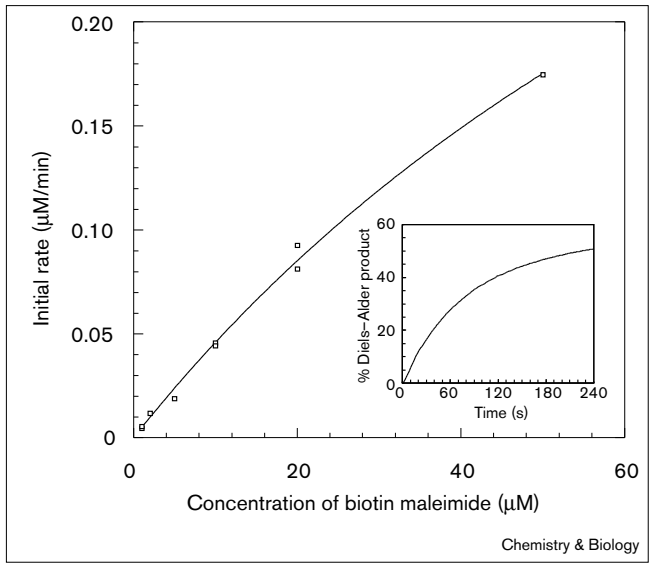
whereas in 39M49 there is a 5 bp helix I and a 4 bp helix III. Both versions accelerated the Diels–Alder reaction almost 2 × faster than their parental sequence with k_{app} values of 142 and 144 $M^{-1}s^{-1}$, respectively. The length of the helices, therefore, does not seem to influence the catalytic performance. For these two sequences, the folding algorithm did not propose other reasonable secondary structures, confirming our assumptions about the structure of the active motif.

Kinetic characterization and salt dependence of the 49-mer ribozyme

We chose the improved 49 nucleotide RNA 39M49 for further characterization. None of the transition metals, Lewis acids and other cofactors present in the reaction mixture were utilized by the catalyst. The Mg^{2+} dependence of the apparent rate constant is shown in Figure 5, where a hyperbolic profile was observed. With longer RNAs, high Mg^{2+} concentrations often lead to aggregation and precipitation; this was not the case with the 49-mer, however. At 80 mM Mg^{2+}, the apparent rate constant was 245 $M^{-1}s^{-1}$. The same rate acceleration was produced with 5 mM Mn^{2+} or 5 mM Mg^{2+}. Replacement of Mg^{2+} with 5 mM Ca^{2+}, however, caused a 65% reduction of activity. The monovalent cations were also exchangeable. The presence of 300 mM Na^+ or 300 mM K^+ gave rate accelerations within 10% of the rate accelerations determined under the selection conditions of 200 mM Na^+ and 100 mm K^+. These findings suggest that catalysis was achieved in a different way from the catalyst reported by Tarasow *et al.* [10] where the presence of both copper ions and covalently bound pyridyl moieties were essential.

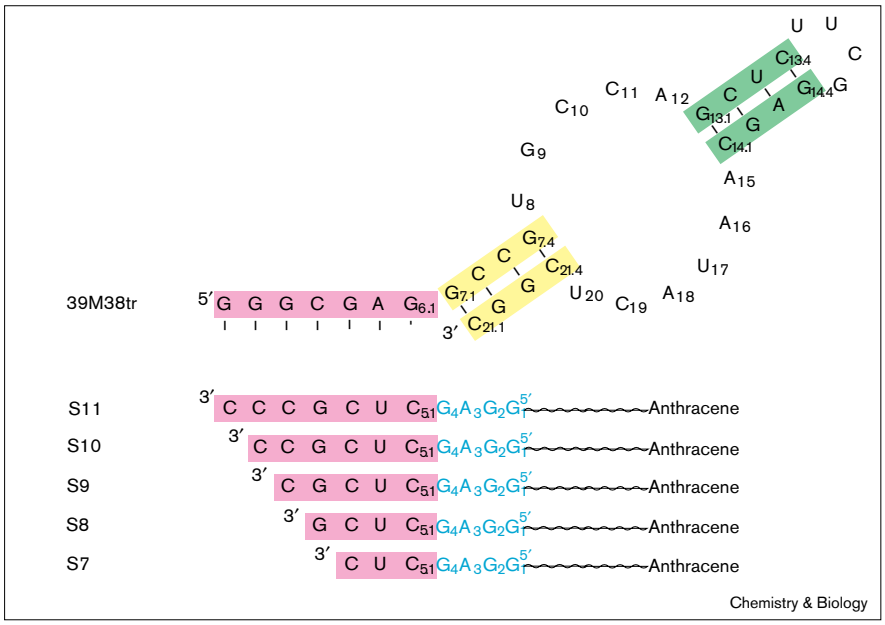
The high reactivity of biotin maleimide prevented a complete kinetic analysis. Quantitative estimation of initial rates at maleimide concentrations higher than 50 µM requires the application of fast kinetic methods, because with 50 µM maleimide 10% of the anthracene has reacted in less than 20 seconds (see inset in Figure 6). Experiments in the range of 1–50 µM biotin maleimide indicate that 39M49 obeys Michaelis–Menten kinetics, with a K_M of 119 ± 24 µM and a k_{cat} of 1.18 ± 0.18 min^{-1} in the presence of 5 mM Mg^{2+} (Figure 6). The k_{cat}/K_M is 167 $M^{-1}s^{-1}$, which is in good

Figure 6



Determination of k_M (using nonlinear least square regression data) and k_{cat} (the Michaelis–Menten first-order rate constant for catalysis) for the self-modifying reaction of the engineered molecule 39M49. Inset, time course for the reaction of 39M49 with 50 µM biotin maleimide in the presence of 5 mM $MgCl_2$.

Figure 7



(a) The proposed secondary structure of the *trans* ribozyme 39M38tr and **(b)** the set of substrate oligonucleotide conjugates. The color scheme is the same as in Figure 4.

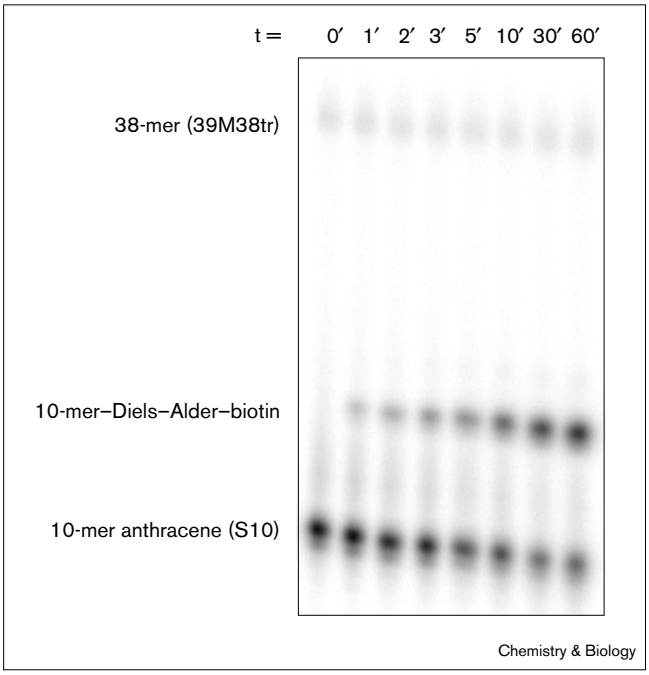
agreement with the apparent second-order rate constants (k_{app}) measured in the assays.

Trans reaction and true catalysis

To study whether individual ribozymes could act as true catalysts, we first screened all clones and minimum structures in the presence of anthracene–PEG–GMP and biotin maleimide. There was, however, no observable rate acceleration under our assay conditions. We then prepared a 38-mer ribozyme (39M38tr) derived from 39M49 by formally cutting loop 1. A series of substrate oligonucleotide–PEG–anthracene conjugates ranging from 7–11 nucleotides in length were prepared, engineered to form double strands with the ribozyme with lengths of 3–7 bp, respectively (Figure 7). Samples were incubated under single-turnover conditions and aliquots withdrawn at the specified time points were analyzed using denaturing polyacrylamide gel electrophoresis (PAGE). At ambient temperature (23°C), 4 bp was the minimum double-strand length required for substrate recognition and efficient catalysis. Figure 8 shows the reaction of S10 (6 bp) with biotin maleimide, catalyzed by 39M38tr under high Mg^{2+} (80 mM) conditions. From the ratio of the bands, k_{app} was calculated to be 208 $M^{-1}s^{-1}$, which corresponds to a 16,000-fold rate acceleration and is only 15% slower than the self-modifying reaction of 39M49.

Ribozyme 39M38tr is capable of catalyzing multiple turnovers. Under substrate-saturating conditions, 1.9 turnovers were measured within 4 h after correction for the uncatalyzed reaction (data not shown). This value is rather low, compared with the k_{cat} values measured for

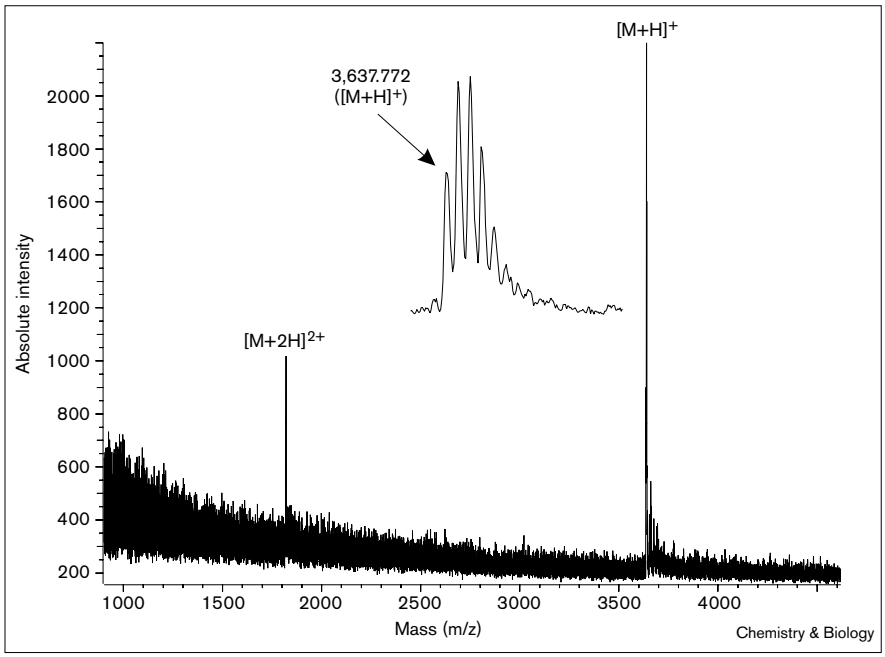
Figure 8



Catalysis of the intermolecular reaction under single-turnover conditions. Reactions were performed in a minimal reaction buffer containing 300 mM NaCl, 80 mM $MgCl_2$ and 30 mM Tris–HCl at pH 7.4 at room temperature. The concentration of the substrate S10 was 1 nM, biotin maleimide 5 μM and ribozyme 39M38tr 0.5 μM. Aliquots were withdrawn at the indicated time points and reactions stopped by mixing with 2 volumes of stop solution (80% formamide, 20 mM EDTA, 100 mM β-mercaptoethanol) and freezing to −20°C. Samples were analyzed on a denaturing 18% polyacrylamide gel and quantified on a phosphorimager.

Figure 9

MALDI–TOF mass spectrometry of the product from the intermolecular reaction of S8 with biotin maleimide catalyzed by 39M38tr. The calculated molecular weight of the biotin–(Diels–Alder product)–PEG8-5′-GGAGCUCG-3′ is 3,637.845 Da ([M+H]$^+$). The inset shows monoisotopic resolution of the [M+H]$^+$ peak.



the self-modifying reaction, and therefore indicates that the chemical step (bond formation) is not rate limiting.

Finally, we investigated the chemical nature of the reaction product using matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) mass spectroscopy. The product of the S8 *trans* reaction was purified under similar conditions to those described in Figure 8, excised from the gel and analyzed. The mass spectrum is shown in Figure 9, and the resulting m/z value is in exact agreement with the calculated value.

Mechanistic investigations

In order to gain insight into the molecular determinants of substrate recognition, we performed inhibition and competition studies using fragments of biotin maleimide (Figure 10). A 100-fold excess of biotin, *N*,*N*′-dibutyrylhydrazine or γ-maleimidobutyric acid, over biotin maleimide, was added to the standard fluorescence assay of the self-modifying reaction catalyzed by 39M49. The reaction appeared to be unaffected, however; apparent rate constants were within 10% of the standard reaction. These results suggest a multipoint recognition of biotin maleimide, as was demonstrated recently for a biotin-binding RNA pseudoknot structure [18].

To investigate the influence of the tether length, we prepared three different versions of substrate oligonucleotide S10, in which the tether consisted of 7, 10 or 16 ethylene glycol units, and monitored the reaction with 39M38tr electrophoretically. The apparent rate constants did not differ significantly from each other (187, 199 and 212 $M^{-1}s^{-1}$, respectively), demonstrating that, in this range, the number of ethylene glycol units is not critical.

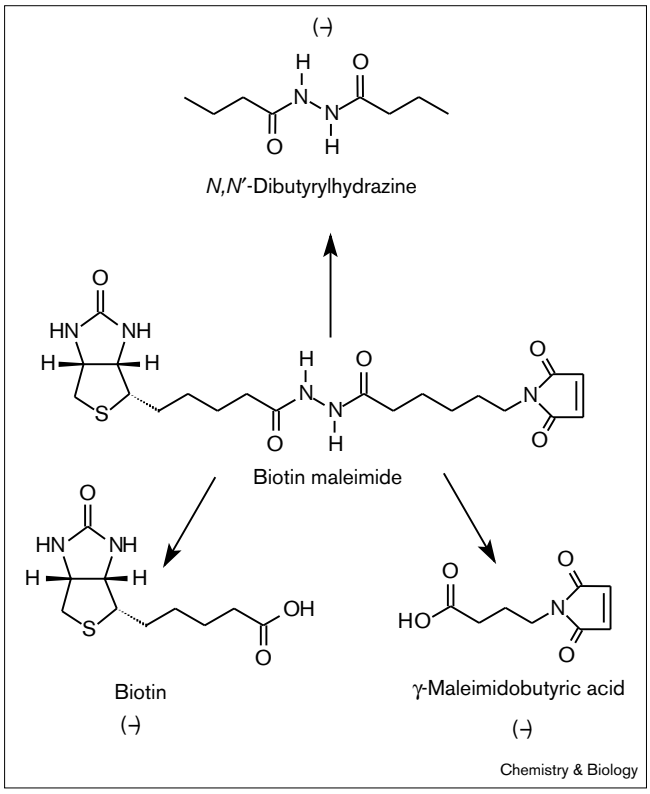
Several known ribozymes have been reported that are able to catalyze both the forward and the reverse reactions (e.g., [19]). To investigate this possibility for the Diels–Alderase catalytic motif, we excised the product bands from the gel shown in Figure 8, and added the desalted conjugates to a reaction mixture containing catalyst 39M38tr in the minimum catalysis buffer under single turnover conditions. No bands with higher electrophoretic mobility were detected, ruling out this possibility (data not shown).

Discussion

The Diels–Alder reaction

We specifically chose the reaction between anthracene and maleimide for a number of reasons. Most importantly, we thought that the completely different overall geometry of reactants and products would facilitate the enrichment of catalysts that are capable of multiple turnovers. Anthracene is planar, in contrast to the 120° angles between the different rings in the Diels–Alder product. A ligand that can bind to anthracene, therefore, should not be able to bind the product except after extensive refolding. Specific RNA sequences might use stacking of anthracene on top of base pairs or base triples, or intercalation for binding, as shown for various RNA aptamers (e.g., [20]). After the reaction, these structures would be unstable.

Figure 10



Inhibition and competition studies with fragments of biotin maleimide. Ribozyme 39M49 (500 nM) was dissolved in minimal reaction buffer (see Figure 8), and the reactions were initiated by adding a 1:100 mixture of biotin maleimide and the respective analogs shown to give final concentrations of 10 μM biotin maleimide and 1 mM analog. The reactions were monitored using fluorescence spectrometry. (–), no effect.

Anthracene is very hydrophobic and, without coupling to PEG, is barely soluble in water. In this study, RNA has found a way to utilize this hydrophobic ring system as a substrate in an enzymatic reaction. It might be that this planar, hydrophobic structure interacts more easily with RNA than do the aliphatic dienes tried in other laboratories. The chemically modified RNA catalysts for the reaction between a similar biotinylated maleimide and an aliphatic diene recently reported by Tarasow *et al.* [10] were much bigger, had more complex structures and had significantly lower rate accelerations.

The ribozymes

We have identified a considerable variety of new ribozymes capable of catalyzing C–C bond formation. The sequence data compiled in Figure 3 show that we have found 13 highly different ribozymes that share a common secondary-structure motif, which can be traced back to 13 different sequences in the ancestral pool. Point mutations inside the sequence families probably stem from misincorporations during the enzymatic amplification steps. The common motif is highly abundant and must have been present in the starting pool in several thousand variants. The fact that we found 35 different active sequences in the 55 analyzed clones indicates that the complexity of the round 10 pool was still rather high, and that many more active species are probably present.

The *trans* version of the Diels–Alder catalyst (39M38tr) is only 38 nucleotides long. Oligonucleotides of this size range have been synthesized by mineral-catalyzed random polymerization of nucleotides under assumed prebiotic conditions [21]. If RNA molecules of this size can solve complex catalytic tasks without base-pair templating of the respective reactants, similar motifs could have been involved in prebiotically relevant steps, such as the synthesis of nucleotides or cofactors [22]. The fact that these motifs have not been found yet, is — in the authors' opinion — solely caused by limitations in the current selection methodology.

It is interesting that the RNA is not selective about cofactors. The ribozyme requires a certain concentration of monovalent and divalent ions, but it can perform its task using a variety of different ions. This property makes the ribozyme functional under very different environmental conditions. In order to increase chances of selecting catalysts, we had added a number of potential facilitators. After the report by Tarasow *et al.* [10], we revised our selection buffer to include Lewis acids and ligands containing pyridyl residues. The sequences isolated in our selection, did not utilize these facilitators, however, indicating that for this particular reaction RNA has found a more efficient solution than general Lewis-acid catalysis.

Although the formation of C–C bonds is central to biochemical pathways, Diels–Alder reactions are not the preferred reaction (nature prefers aldol reactions). Nevertheless, proteins are capable of catalyzing Diels–Alder cycloadditions. Several catalytic antibodies raised against the respective transition-state analogs exhibited Diels–Alderase activities and demonstrated remarkable enantioselectivity and mechanistic preferences [6,7]. The RNA motif described here compares quite well with these antibodies — the rate acceleration is higher and the molecules are considerably smaller. In terms of rate acceleration, the ribozyme catalysts described here belong to the most efficient macromolecular Diels–Alderases known.

The selection methodology

Although direct selection had been applied to numerous self-modifying reactions for several years, its extension to reactions involving two nonRNA reactants succeeded only recently [10,23]. The attachment of a potential reactant via a flexible polymeric tether allows the ribozyme chemistry to be extended to numerous chemical reactions that were previously inaccessible [24,25]. In this study, this methodology

has for the first time allowed a small RNA motif capable of catalyzing C–C bond formation to be isolated, thereby expanding our knowledge about RNA catalysis.

Although we succeeded in expanding the chemistry of RNA catalysis, our selected sequences fail to accelerate the reaction between the two free reactants. We assume that our ribozyme has only a low-affinity binding site for anthracene, and that an efficient reaction acceleration requires a tremendous increase in the local anthracene concentration, which is acheived by tethering anthracene to the RNA conjugate library during the selection. A conservative estimate using an extended spacer length of ~45 Å and neglecting the excluded volume, which is due to the occupation of space by RNA, suggests that the lower boundary of the local anthracene concentration in the vicinity of a potential active site is ~5 mM, which is a 10,000-fold increase over the actual concentration. The applied selection pressure, therefore, does not induce the selection of species containing a high-affinity binding site for anthracene, a requirement for recognizing free anthracene at nanomolar or micromolar concentrations.

Significance

The formation of carbon–carbon bonds is an essential reaction in the metabolism of a hypothetical RNA world, which might have been a precursor to extant life. Using direct selection with linker-coupled reactants, we have isolated a set of new ribozymes that catalyze the formation of carbon–carbon bonds by Diels–Alder reaction. Comprehensive sequence analysis allowed us to condense the ~150-nucleotide-long RNAs to a 38-mer ribozyme that is not only faster than most full-length sequences but also acts as a true enzyme in an intermolecular way. The enzyme exhibits a remarkably broad tolerance for different metal ions, making it functional under very different conditions.

In this system, catalysis of a complex reaction that involves the creation of two bonds is achieved by a small oligoribonucleotide that does not require direct positioning of the two reactants by base pairing. The catalytic activity described here reveals the potential of small ribozymes for performing comprehensive organic transformations, which is essential not only for recreating prebiotic reaction pathways but also for the application of RNA catalysts in bioorganic synthesis.

Materials and methods

Materials

T7 RNA polymerase was purchased from Stratagene, [α-^{32}P]-CTP from Amersham, DNase I (RNase free) and SuperScript II RNase H$^-$ reverse transcriptase from Gibco, Taq polymerase and bovine serum albumin from Boehringer Mannheim. Unlabeled NTPs were obtained from Boehringer Mannheim and dNTPs from MBI Fermentas. Anthracene–PEG–guanosine conjugates (initiator nucleotides) were synthesized as described previously [11,12]. Primers and synthetic oligonucleotides were either synthesized on an Applied Biosystems 394 synthesizer or purchased from MWG Biotech. Biotin maleimide, streptavidin agarose and γ-maleimidobutyric acid were purchased from Sigma, biotin from Aldrich, 2,3-di(4-pyridyl)-2,3-butanediol from Fluka, and H-His–Lys-OH from Bachem. Spin filters (Ultrafree-MC 0.45 μm) were from Millipore. Cloning was carried out using the TOPO TA Cloning kit (Invitrogen). Sequencing was carried out using the dideoxy method with Thermo Sequenase (Amersham) on an LI-COR DNA Sequencer 4000 (LI-COR Biotech). UV spectra were recorded with a Shimadzu UV-160A and fluorescence measurements were done with the Luminescence Spectrometer LS50B from Perkin Elmer. *N,N′*-Dibutyrylhydrazine was synthesized as described previously [26].

Preparation of the RNA pool

A 157-mer DNA (5′-GGAGCTCAGCCTTCACTGC–N_{120}–GGCACCACGGTCGGATCC-3′) with a random insert of 120 nucleotides was synthesized, where N represents a mixture of A,C,G and T due to a mole ratio of 3:3:2:2 A:C:G:T phosphoramidites in DNA synthesis. To increase product yield, capping was omitted during synthesis of the randomized part. The product was deprotected in 33% ammonia, lyophilized and desalted. After purification on an 6% denaturing polyacrylamide gel, the band was visualized by UV shadowing, excised and eluted with 0.3 M NaAc pH = 5.5 and precipitated with ethanol. About 15% of the DNA was PCR amplifiable. 1.3 nmol of the DNA was PCR-amplified in a total volume of 20 ml PCR mix (20 mM Tris–HCl pH 8.4, 50 mM KCl, 4 mM $MgCl_2$, 0.2 mM dNTPs, primer A (5′-TCTAATACGACTCACTATAGGAGCTCAGCCTTCACTGC-3′; 5 μM) and primer B (5′-GTGGATCCGACCGTGGTGCC-3′; 7.5 μM)) in eight cycles (temperature cycle conditions: 94°C, 2 min; 50°C, 6 min; 72°C, 10 min). The PCR amplified DNA was extracted with phenol/chloroform and precipitated with ethanol. For *in vitro* transcription, 950 pmol of PCR-amplified DNA was used in a 3 ml transcription reaction (80 mM HEPES pH 7.5, 1 mM spermidine, 22 mM $MgCl_2$, 10 mM dithiothreitol, 240 μg/ml bovine serum albumin, 4 mM NTPs, 2 mM initiator nucleotide, 3 μM [α-^{32}P]-CTP (0.2 μCi/μl), 1 U/μl T7 RNA polymerase) and transcribed at 37°C for 4 h. 300 U DNase I was added and incubation was continued for 40 min. RNA was purified on a 10% denaturing polyacrylamide gel, visualized, excised, eluted and precipitated 3 × as described above and used for selections. The fraction of RNA conjugated with anthracene–PEG at the 5′-end was found to be about 50% by high performance liquid chromatography [12].

In vitro *selection*

In vitro selection was performed with an initial pool of approximately 2×10^{14} sequences using 50 copies of each molecule in the first cycle. The RNA library was dissolved to 1 μM concentration in 200 mM NaCl, 100 mM KCl, 30 mM Tris–HCl, pH 7.4 and heated to 90°C for 5 min to denature the RNA. The reaction mixture was cooled down to room temperature for 20 min. All divalent metal ions and potential cofactors were added to give a final concentrations of 5 mM $MgCl_2$; 5 mM $CaCl_2$; 5 μM each of $AlCl_3$, $CoCl_2$, $CuCl_2$, $MnCl_2$, $ZnCl_2$, 2,3-di(4-pyridyl)-2,3-butanediol, H-His–Lys-OH. The reaction was initiated by the addition of biotin maleimide in dimethyl sulfoxide (DMSO) to give a concentration of 25 μM biotin maleimide and 2% DMSO and incubated at 25°C for 1 h. Selection stringency was gradually increased by decreasing the biotin maleimide concentration to 2.5 μM and the incubation time to 1 min (see Figure 2). The RNA was ethanol precipitated twice, resuspended in streptavidin-binding buffer (1 M NaCl, 10 mM HEPES pH 7.2, 5 mM EDTA), denatured at 90°C for 5 min and cooled on ice following incubation for 45 min with streptavidin agarose (washed twice with streptavidin-binding buffer including 2 mg/ml tRNA to saturate possible unspecific RNA-binding sites). The sample was transferred to a spin filter and washed 20 × with denaturing buffer (8 M urea, 0.1 M Tris–HCl pH 7.4) and twice with water. RNAs linked to the streptavidin resin were directly reverse transcribed according to the manufacturers instructions with 0.3 μg/μl bovine serum albumin at 55°C, PCR amplified and T7 transcribed. Beginning with round 3, in every other round a preselection step against streptavidin binders was introduced: after the initial denaturation step, RNA was cooled on ice and incubated with streptavidin agarose for 45 min. The supernatant was heated to 90°C

for 5 min and cooled to room temperature for 20 min again, followed by addition of the divalent metal ions and so on.

Apparent rate constants k_{app} were calculated from the percentage immobilized $\%_{immo}$ by:

$$k_{app} = \frac{1}{t([BM]_0 - [A-RNA]_0)} \times \ln \frac{[A-RNA]_0[BM]_t}{[A-RNA]_t[BM]_0}$$

in which $[BM]_0$ and $[A–RNA]_0$ are the initial concentrations of biotin maleimide and anthracene–RNA, respectively, and the concentrations of anthracene–RNA and biotin maleimide at the time t $[A–RNA]_t$ and $[BM]_t$ were calculated from:

$[A–RNA]_t = (100 - \%_{immo})/100\ ([A–RNA]_0)$ and

$[BM]_t = [BM]_0 - [A–RNA]_0 + [A–RNA]_t$.

Pool DNA obtained after round 10 was cloned and sequenced following the manufacturer's instructions.

Kinetic analysis

All kinetic experiments were carried out at ribozyme concentration of 0.5 μM in selection buffer at room temperature unless otherwise stated. Spectroscopic assay of the anthracene fluorescence decay was performed with 50 μM biotin maleimide or in the range 1–50 μM for Michaelis–Menten kinetics upon excitation at 253 nm measuring emission at 419 nm. Kinetic measurements of all selected ribozymes, of ribozymes from deletion analysis, for inhibition studies and for investigating salt dependence were done by fluorescence spectroscopy. Values of k_{app} were extrapolated either graphically or by fitting the experimental values to a standard second-order rate equation.

Intermolecular reactions under multiple turnover conditions were performed similar to those described in Figure 8, except that substrate concentration was 10 μM and biotin maleimide 50 μM. Monodisperse substrate oligonucleotides S10 were prepared by using the respective monodisperse initiator nucleotides derived from polydisperse mixtures by reversed phase HPLC.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (grants # Ja 794/1 and SFB 344/B10 to A.J.). We thank E. Nordhoff for mass spectrometric analysis, M. Ziegler and G. Buchlow for assistance with sequencing and fluorescence spectrometry, and V.A. Erdmann for support. We appreciate helpful suggestions by F. Hausch, C. Frauendorf, and T. Ruppert.

References

1. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. (1995). Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* **64**, 763-797.
2. Osborne, S.E. & Ellington, A.D. (1997). Nucleic acid selection and the challenge of combinatorial chemistry. *Chem. Rev.* **97**, 349-370.
3. Robertson, M.P. & Ellington, A.D. (1998). How to make a nucleotide. *Nature* **395**, 223-225.
4. Breaker, R.R. (1997). *In vitro* selection of catalytic polynucleotides. *Chem. Rev.* **97,** 371-390.
5. Famulok, M. & Jenne, A. (1998). Oligonucleotide libraries – *variatio delectat. Curr. Opin. Chem. Biol.* **2**, 320-327.
6. Hilvert, D., Hill, K.W., Nared, K.D. & Auditor, M.-T.M. (1989). Antibody catalysis of a Diels–Alder reaction. *J. Am. Chem. Soc.* **111**, 9261-9262.
7. Gouverneur, V.E., Houk, K.N., de Pascual-Teresa, B., Beno, B., Janda, K.D. & Lerner, R.A. (1993). Control of the exo and endo pathways of the Diels–Alder reaction by antibody catalysis. *Science* **262**, 204-208.
8. Kang, J., Santamaria, J., Hilmersson, G. & Rebek, J.J. (1998). Self-assembled molecular capsule catalyzes a Diels–Alder reaction. *J. Am. Chem. Soc.* **120**, 7389-7390.
9. Morris, K.N., Tarasow, T.M., Julin, C.M., Simons, S.L., Hilvert, D. & Gold, L. (1994). Enrichment for RNA molecules that bind a Diels–Alder transition state analog. *Proc. Natl Acad. Sci. USA* **91**, 13028-13032.
10. Tarasow, T.M., Tarasow, S.L. & Eaton, B.E. (1997). RNA-catalysed carbon-carbon bond formation. *Nature* **389**, 54-57.
11. Seelig, B. & Jäschke, A. (1997). Site-specific modification of enzymatically synthesized RNA: transcription initiation and Diels–Alder reaction. *Tetrahedron Lett.* **38**, 7729-7732.
12. Seelig, B. & Jäschke, A. (1999). Ternary conjugates of guanosine monophosphate as initiator nucleotides for the enzymatic synthesis of 5'-modified RNAs. *Bioconjugate Chem.* **10**, in press.
13. Rideout, D.C. & Breslow, R. (1980). Hydrophobic acceleration of Diels–Alder reactions. *J. Am. Chem. Soc.* **102**, 7816-7817.
14. Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881-10890.
15. Matzura, O. & Wennborg, A. (1996). RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Computer Applications in the Biosciences (CABIOS)* **12**, 247-249.
16. Jaeger, J.A., Turner , D.H. & Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc. Natl Acad. Sci. USA* **86**, 7706-7710.
17. Walter, A.E., *et al.,* & Zuker, M. (1994). Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl Acad. Sci. USA* **91**, 9218-9222.
18. Wilson, C., Nix, J. & Szostak, J.W. (1998). Functional requirements for specific ligand recognition by a biotin-binding RNA pseudoknot. *Biochemistry* **37**, 14410-14419.
19. Chowrira, B.M., Berzal-Herranz, A. & Burke, J.M. (1993). Ionic requirements for RNA binding, cleavage and ligation by the hairpin ribozyme. *Biochemistry* **32**, 1088-1095.
20. 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.
21. Ferris, J.P., Hill, A., Jr., Liu, R. & Orgel, L.E. (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59-61.
22. Unrau, P.J. & Bartel, D.P. (1998). RNA-catalyzed nucleotide synthesis. *Nature* **395**, 260-263.
23. Wiegand, T.W., Janssen, R.C. & Eaton, B.E. (1997). Selection of RNA amide synthases. *Chem. Biol.* **4**, 675-683.
24. Hausch, F. & Jäschke, A. (1997). Libraries of multifunctional RNA conjugates for the selection of new RNA catalysts. *Bioconjugate Chem.* **8**, 885-890.
25. Frauendorf, C. & Jäschke, A. (1998). Catalysis of organic reactions by RNA. *Angew. Chem. Int. Ed.* **37**, 1378-1381.
26. Wu, D.Y. & Herbst, R.M. (1952). Substituted phenyltetrazoles, alkoxynitrophenyle tetrazoles and alkoxyaminophenyl tetrazoles. *J. Org. Chem.* **17**, 216-227.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.