

RNA Funktionskontrolle durch strukturelles Design

Inauguraldissertation

zur

Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Ernst-Moritz-Arndt-Universität Greifswald

Vorgelegt von Sonja Petkovic

Geboren am 01.07.1982

in Hannover

Greifswald, den 31.03.2014

Dekan: Prof. Dr. Klaus Fesser

1. Gutachter: Prof. Dr. Sabine Müller
2. Gutachter: Prof. Dr. Christian Hammann

Tag der Promotion: 01. Juli 2014

Inhaltsverzeichnis

Inhaltsverzeichnis	S. i
1. Einleitung	S. 1
2. Zielstellung	S. 4
3. Zusammenfassende Darstellung und Diskussion der Ergebnisse in zwei Teilen	S. 5
3.1 Teil I: Analysen von CRZ-2 und seinem finalen Spalt-Produkt	S. 7
3.1.1 Analysen der Spalt-Produkte und Nachweis zyklischer RNA	S. 7
3.1.2 Analyse der Multimere	S. 13
3.2 Fazit Ergebnis-Teil I	S. 17
3.3 Teil II: Ergebnisse der bioinformatisch entworfenen Test-Sequenzen	S. 18
3.3.1 Bioinformatischer Entwurf von Hairpin-Ribozym-basierten Test-Sequenzen	S. 18
3.3.2 Vergleich der vorhergesagten Eigenschaften mit experimentellen Ergebnissen	S. 19
3.3.2.1 Übereinstimmungen	S. 19
3.3.2.2 Unterschiede	S. 21
3.4 Struktur-Funktions-Beziehungen und Ausblick	S. 22
4. Literaturverzeichnis	S. 26

1. Einleitung

Diese Arbeit behandelt die Analyse und Funktions-Programmierung von selbst-prozessierenden RNA-Molekülen. Selbst- (oder Auto-) Prozessierung sei definiert als eine kovalente spezifische Modifikation an einem Makromolekül, die das Makromolekül an sich selbst durchführt (hier die reversible Bildung einer Phosphodiesterbindung). In der Natur und *in vitro* lassen sich sowohl RNA- als auch Protein-basierte Selbst-Prozessierungsreaktionen finden (1). Dazu zählen bei Proteinen u.a. die Aktivierung des vakuolen Prozessierungsenzyms aus *Arabidopsis* (2) und die Autophoshorylierung von Tyrosinkinasen, wie z.B. dem Insulinrezeptor (3,4). Selbst-Prozessierung bei RNA erfolgt *in vivo* beim autokatalytischen RNA-Spleißen z.B. in *Tetrahymena* (5) oder wird *in vitro* zur Herstellung von zyklischen RNA Molekülen zum Beispiel für RNA-Interferenz genutzt (6).

RNA kann in biologischen Prozessen synthetisiert, hydrolysiert, verzweigt, zirkularisiert (siehe Manuskript 4), multimerisiert, verkürzt, translatiert, editiert und, wie hier vorgestellt, selbst-prozessiert werden. Diese Vielfältigkeit spiegelt die Komplexität von Prozessen mit und an RNA aus Zellen aller heute existierenden Lebewesen wider. Zudem birgt die Vielfältigkeit auch ein hohes Potenzial der RNA für die Entwicklung präbiotischen Lebens im Sinne der RNA-Welt-Hypothese (7) und verdeutlicht die Notwendigkeit und Faszination für die Analyse selbst-prozessierender RNA. Zyklische RNA im Besonderen wurde schon 1989 von T. Diener als Relikt einer möglichen prä-zellularen Welt diskutiert (8) und kürzlich von Ma *et al.* als mögliches Ur-Chromosom vorgeschlagen (9). Ma *et al.* beschrieben wahrscheinlichkeitsbasierte Simulationen, wobei das Ur-Chromosom je nach vorliegender Struktur entweder katalytisch aktiv sein kann und sich selbst in einzelne Gene spaltet, oder ausschließlich als Informationsträger fungiert und beispielsweise für ein Replikase-Ribozym zur RNA-Replikation kodiert (9). In heute vorkommenden eukaryotischen Zellen kennt man zyklische RNAs aus Spleiß-Prozessen, wobei es zur Ausbildung eines RNA-Lariats kommt (10). Ebenso führt fehlerhaftes, inverses, alternatives Spleißen und fehlerhaftes Auflösen der Lariat-Struktur zu RNA-Ringen aus Exons (11,12), Introns (13,14) oder beidem (12).

Seit 2013 häuften sich Berichte, dass zyklische RNA nicht nur Produkte fehlerhaften Spleißens sind oder ein Reservoir zur Ribonukleotid-Rückgewinnung darstellen, sondern Funktionen in der eukaryotischen Zelle haben. Zu den Funktionen gehören die posttranskriptionale Regulation mit zirkulärer RNA als Antagonist von mikro-RNA oder als

alternativer Bindungspartner für RNA-Bindeproteine; zirkuläre RNA wird außerdem translatiert und in Verbindung mit Krankheiten wie Alzheimer gebracht (13,15-22). Zudem kommt zyklische RNA im Hepatitis Delta Virus (23), in Viroiden (24) und Satelliten-RNAs (25) vor.

Die im *Tobacco Ringspot Virus* gefundene zirkuläre Satelliten-RNA beider Polaritäten trägt die Eigenschaft, sich selbst spalten zu können (26). Die Spalt-Motive wurden untersucht und jeweils Minimal-Konsensus-Sequenzen gefunden, die für die Spalt-Aktivität nötig sind (26). Für *in vitro* Anwendungen wurden RNAs mit diesen Spalt-Aktivitäten modifiziert, umprogrammiert und so „echte“ Biokatalysatoren erhalten, die sich durch multiple Prozessivität auszeichnen, das *Hammerhead*- und *Hairpin*-Ribozym (27,28)(s. Manuscript 1). So verhält es sich auch mit der in dieser Arbeit verwendeten, früher im Arbeitskreis von Prof. Sabine Müller entworfenen, *cis*- und *transaktiven* katalytischen RNA: *cyclic ribozyme 2*, kurz: CRZ-2. Diese Hairpin-Ribozym-Variante ist am 5'-Ende verlängert, und Helix 1 wurde mit einem Loop aus vier Basen (Sequenz zufällig gewählt) geschlossen. Helix 4 ist mit dem ultrastabilen Tetraloop 5'-UUCG-3' geschlossen (29,30), und die Verknüpfung der zwei *junctions* erfolgte mittels Pentacytidin (31) (Manuscript 2, Abb. 1a). Diese katalytische RNA kann in zwei Vorzugs-Konformationen vorliegen und ist somit eine bistabile RNA. Sie hat dadurch die Fähigkeit, zwei Spalt-Reaktionen zu vollziehen und so zwei vormals inhärente Sequenzabschnitte, welche den 5'- und den 3'-Terminus bildeten, freizusetzen. Eine Erhöhung der Magnesium-Konzentration kann Liggationsreaktionen begünstigen (32) (s. Manuscript 2, Abb. 1a, b). Dies resultiert in RNA-Monomeren und Oligomeren, die sowohl zyklisch als auch linear sein können (33). Ein RNA-Monomer sei hier eine CRZ-2-Variante der Länge 83-, 91-, 92-, 94-, 97mer, (mögliche nicht-zyklisierte Spalt-Produkte und deren DNA-Äquivalente), zyklisches 83mer, 103-, oder 105mer (mögliche Transkripte).

Die Katalyse bedient sich folgender Strategien: Zu Beginn kommt es zur präzisen Substratorientierung und zu koaxialem *Stacking* und anschließend zu einer tertiären Interaktion, durch Loop A- und Loop B-Kontakte (32). Folgend werden diverse Interaktionen im *Hairpin*-Ribozym beschrieben. Die Nummerierungen der beteiligten Nukleobasen beziehen sich auf die in den Referenzen angegebenen Positionen. An den Interaktionen sind beteiligt:

- i) kanonische Basepaare innerhalb der Helices für den allgemeinen Aufbau des Ribozyms,

- ii) nichtkanonische Basenpaare innerhalb der Loop-Regionen,
- iii) ein Ribose-Zipper-Motiv zwischen den N3 Atomen von A10 und A24, sowie den 2'- OH-Gruppen von A10, G11, A24 und C25, (34,35)
- iv) Wasserstoffbrücken zwischen U 42 und G11, U12, A22 und A23, (34)
- v) eine Bindungstasche innerhalb der Loopstruktur B, bestehend aus A38 und A26 zur Fixierung von G+1 (36) und
- vi) ein Interdomänen Watson-Crick-Basen-Paar zwischen G+1 und C25 (37).

Drei Mechanismen kommen wenigstens in Frage: i) allgemeine Säure-Base-Katalyse vermittelt durch die Nukleobasen Adenin 38 und Guanin 8 (38,39), ii) Konformationseffekte zur Unterstützung der *in-line* Geometrie (40), und iii) elektrostatische Stabilisierung des Übergangszustandes (41). Für die Strategie i) ist eine Verschiebung von pKa-Werten der funktionellen Gruppen einzelner Nukleobasen in den hier katalytisch relevanten pH-Bereich um 7 nötig, z.B. über Wasserstoffbrückennetzwerke (42). Es kommt zu einer Reduktion der Aktivierungsenergie und die Reaktionsgeschwindigkeit wird um den Faktor 10^5 erhöht (43). Die Spaltung verläuft über einen assoziativen S_N2 -artigen Mechanismus mit einem *in-line* Angriff des 2'-Sauerstoffs auf das verbrückende Phosphoratom (44). Angreifende und austretende Gruppe stehen in apikalen Positionen der intermediär gebildeten trigonalen Bipyramide (s. Manuskript 2, Abb. 2). Es entsteht ein 2',3'-Zyklophosphat und eine Hydroxygruppe am 5'-Terminus (26). Der Mechanismus ist bei den hier verwendeten 37 °C und neutralem pH-Wert reversibel. Mikro- oder auch mikroskopische Reversibilität durch die Ribozymstruktur gewährleistet Regioselektivität, so dass es nahezu ausschließlich zur Bildung von 3',5'-Phosphodiestern kommt.

2. Zielstellung:

Initial wurde das Hairpin-Ribozym-basierte System CRZ-2 verwendet. Die Produkte der Selbst-Prozessierung nach Spalt- und Ligationsreaktion von CRZ-2 sollten untersucht werden. Dafür war es nötig zunächst den Ablauf der Spalt-Kaskade nachzuweisen und die Produkte zu identifizieren. Inter- und intramolekulare Ribozym-Reaktionen sollten mit CRZ-2 und dem entsprechenden finalen Spalt-Produkt, dem linearen 83mer (l-83mer), durchgeführt und Oligomere und zyklisierte RNAs nachgewiesen werden. Dazu mussten Analyse-Strategien gefunden und ggf. kombiniert werden. Das Wissen um die Art der Reaktionsprodukte von CRZ-2 und dem l-83mer ermöglichen es, diese beiden RNAs als Referenz-Systeme zu benutzen. Diese Referenzen und die neu gefundenen Analyse-Strategien sollten für Untersuchungen weiterer selbst-prozessierender Varianten des Hairpin-Ribozyms genutzt werden. Eine Möglichkeit experimentell neue Varianten einer funktionellen RNA zu finden, ist das Verfahren SELEX (kurz für: *Systematic Evolution of Ligands by Exponential Enrichment*) (45). Dies ist im Arbeitskreis von Prof. Müller durch Dr. T. Marschall erfolgreich angewendet worden um spezifische Aptamere für den medizinisch relevanten Plättchenfaktor 4 zu erhalten (s. Manuskript 5).

In dieser Arbeit erfolgte die Entwicklung neuer Varianten bioinformatisch in Kooperation mit der Arbeitsgruppe von Prof. Ivo Hofacker in Österreich, Universität Wien, durch den Promotionsstudenten Stefan Badelt. Dabei wurde ein wahrscheinlichkeitsbasierter Entwurf (*probability based design*, kurz: PBD) für RNA-Sekundärstrukturen mittels Programm *Switch.pl* (46) aus dem *Vienna RNA package* 2.0 benutzt (47) (s. auch Ergebnisteil II). Die für die katalytische Aktivität der Hairpin-Ribozym-Varianten essentiellen Basenfolgen in den Loops (48-50) wurden beibehalten. Alle Test-Systeme sollten bistabil sein und somit die für CRZ-2 übliche Spalt-Kaskade benutzen. Unterschieden sollen sich die Ribozym-Varianten in Bezug auf ihr Ligationsverhalten: Es sollten Ribozyme gefunden werden, die vornehmlich Monomere zirkularisieren, und den Varianten gegenübergestellt werden, die Dimere zirkularisieren. Somit mussten die RNA-Sequenzen synthetisiert, Ribozym-Reaktionen durchgeführt, die Spalt-Produkte für den Nachweis der Spalt-Kaskade identifiziert, Zykлизierungen nachgewiesen und Oligo- oder sogar Multimerisierungen gezeigt werden. Die Ergebnisse sollten mit denen vom Referenz-System CRZ-2 und seinem finalen Spalt-Produkt verglichen werden. Sequenz-Alignments der Test-Systeme zu CRZ-2 sollten dazu dienen,

Rückschlüsse auf die Funktionen einzelner Basen zu ziehen. In Bezug auf die Biochemie der Ribozyme sollte die Notwendigkeit der An- oder Abwesenheit bestimmter Nukleobasen ein initiales Verständnis für die Programmierung der Hairpin-Ribozym-basierten RNAs in Richtung Monomer-Zyklisierung oder Dimer-Zyklisierung liefern. Für zukünftige bioinformatisch gestützte Entwicklungen von Hairpin-Ribozym-Varianten sollte eine erste Evaluierung der verwendeten Methode vorgenommen und ggf. neue Bedingungen für künftige Ribozym-Sequenz-Entwürfe vorgeschlagen werden.

3. Zusammenfassende Darstellung und Diskussion der Ergebnisse in zwei Teilen

Im Folgenden werden Ergebnisse von Spalt-Reaktionen und Ligationsreaktionen dargestellt (je zwei Stunden, 37 °C, 10 mM Tris-HCl, pH 7,0 und 10 mM Mg²⁺ bei der Spalt-Reaktion; einzige Änderung: 50 mM Mg²⁺ bei der Ligationsreaktion). Folgten diese Reaktionen aufeinander, wird der kürzere Begriff Ribozym-Reaktion verwendet.

Gemeinsam haben das Referenz-System CRZ-2 und die Test-Systeme, dass sowohl Spalt-Reaktionen bei Transkripten (103- oder 105mer) (Abb. 1a), intermediären Spalt-Produkten (Abb. 1b), als auch Oligomeren zu kürzeren Oligo- oder Monomeren stattfinden (nicht dargestellt). Zudem ist es möglich dass sich die Termini der Oligomere abspalten (nicht dargestellt). Intermolekulare Ligationen zwischen Spalt-Produkten (Abb. 1d, e und f) und / oder Oligomeren sind ebenso möglich. Die intramolekulare Ligation am Monomer ist ausschließlich dem finalen Spalt-Produkt, dem linearen 83mer, möglich (Abb. 1c). Das lineare 83mer von CRZ-2 ist im Gegensatz zu denen der Test-Systeme isoliert und separat untersucht worden. Dabei sind sowohl Ribozym-Reaktionen, als auch ausschließlich Ligationsreaktionen durchgeführt worden. Das Ergebnis ist wie erwartet identisch, da keine abspaltbaren Termini vorliegen, sodass auf die vorausgehende Spalt-Reaktion bei diesem System verzichtet wurde. Ohne Spezifikation bezieht sich der Begriff 83mer auf das des Systems CRZ-2.

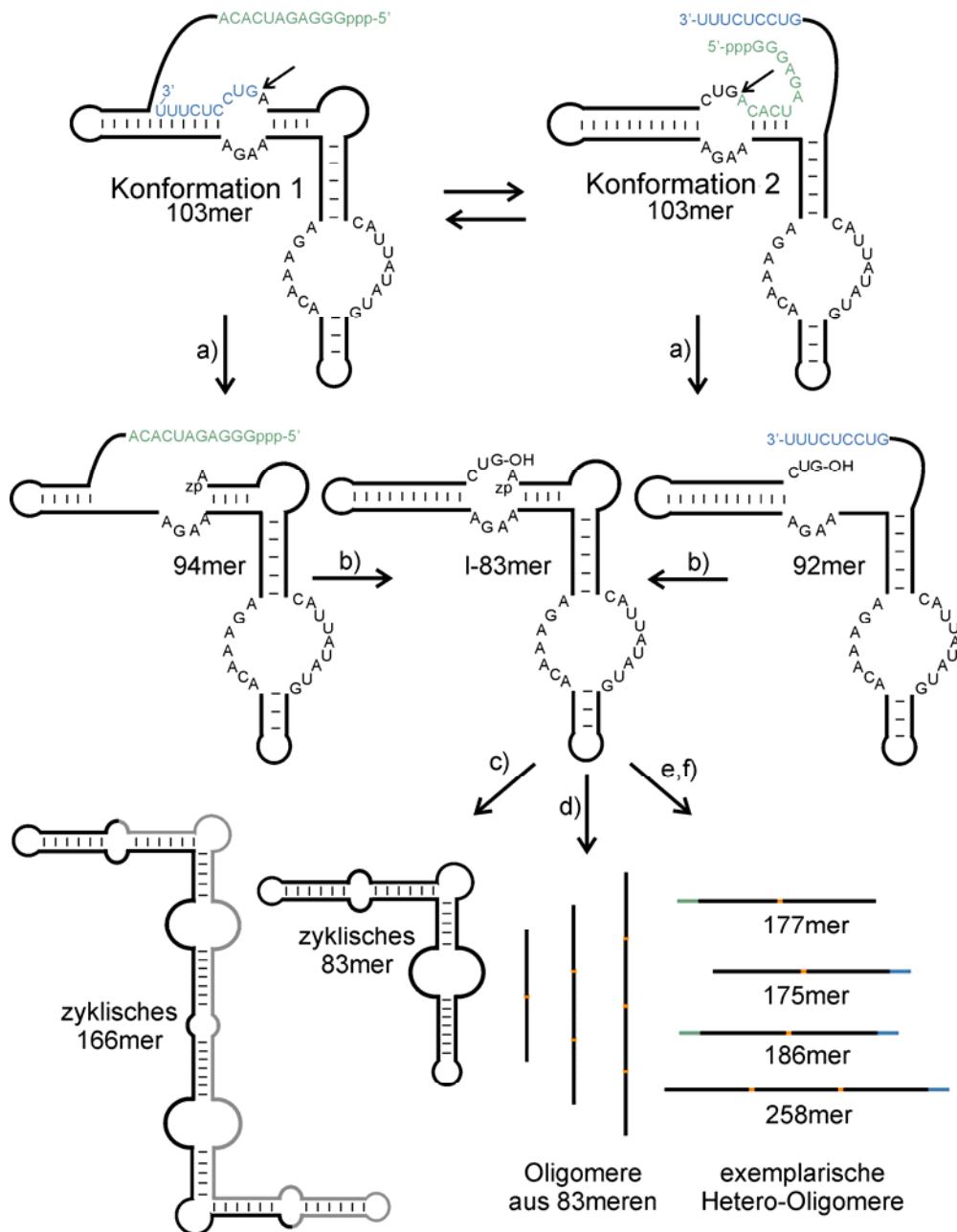


Abb. 1: Überblick zu den möglichen Selbst-Prozessierungen von CRZ-2, den Test-Systemen und den Spalt-Produkten (abweichende Sequenz-Längen bei Test-System 3 und 4 zugunsten der Übersichtlichkeit nicht dargestellt). Zuoberst CRZ-2 in den zwei Vorzugs-Konformationen, die Spaltstelle ist jeweils in Loop A mit einem Pfeil kennzeichnet. a) Jeweils erste Abspaltung eines Terminus (farbig), welche zum jeweiligen intermediären Spalt-Produkt führt, zp bezeichnet das 2',3'-Zyklophosphat. b) zweites Spaltereignis, welches das finale Spalt-Produkt hervorbringt. c) Intramolekulare Ligationen des I-83mers, schematische Sekundärstrukturen des Monomer und Dimer. d) Exemplarische schematische Homo-Oligomere des I-83mers gezeigt. Enthaltende Anzahl Monomere durch orangefarbene Begrenzung dargestellt. e) Mögliche Hetero-Dimere mit und f) ohne I-83mer. Exemplarisches Trimer gezeigt, stellvertretend auch für etwaige Multimere.

3.1 Teil I: Analysen von CRZ-2 und seinem finalen Spalt-Produkt

3.1.1 Analysen der Spalt-Produkte und Nachweis von zyklischer RNA

Die Produkte der Ribozym-Reaktion von CRZ-2 wurden gelelektrophoretisch untersucht und das Migrationsverhalten der RNAs mit einem kommerziell erhältlichen RNA Größenstandard abgeglichen (für Informationen zu den Experimenten s. Manuskript 2 und 3). Das triphosphattragende *full-length* Transkript mit einer Länge von 103 Basen bewegte sich erwartungsgemäß etwas schneller durch das denaturierende Polyacrylamid-Gel als die unmodifizierte Marker-RNA des Größenstandards mit 100 Basen Länge (s. Manuskript 2, Abb. 3).

Bestätigt wurde diese Beobachtung durch eine Angleichung der Laufgeschwindigkeit der transkribierten RNA durch Abspaltung des Triphosphates mittels alkalischer Phosphatase (Daten nicht gezeigt).

Im unteren Gelbereich sind zwei statt drei Banden zu erkennen (Manuskript 2, Abb. 3a, b). Bei der genannten Spalt-Kaskade entstehen nach Abspaltung eines RNA-Segments entweder am 3'- oder am 5'-Ende der RNA je ein Intermediat, das 92mer oder das 94mer. Diese können nach Abspaltung des jeweils anderen Substrates in das finale Spalt-Produkt, das I-83mer übergehen (s. Manuskript 2, Abb. 1b; 3a, b). Bei der oberen Bande handelt es sich um das 92mer im Gemisch mit dem 94mer. Diese RNAs sind kaum elektrophoretisch zu trennen, da das theoretisch höher laufende 94mer das Triphosphat trägt und somit ebenso schnell durch das Gel migriert, wie das kürzere Spalt-Kaskaden-Intermediat. Die Annahme, dass es sich um das Gemisch aus 92- und 94mer handelt, wird wie folgt belegt:

- i) Die RNAs aus dem Gemisch lassen sich ausschließlich in ein kürzeres Produkt spalten, das I-83mer.
- ii) Bei Ribozym-Reaktionen entstehen längere RNAs stets als Doppelbanden bzw. Mehrfachbanden, was eine Beteiligung von unterschiedlichen Monomeren bei der Multimerknüpfung bedeuten kann.
- iii) Die relativ einfache Kinetik wie sie beim linearen 83mer zu finden ist, ist nicht mehr anwendbar. Bei dem 92/94mer handelt es sich um zwei RNAs, die nach Ribozym-Reaktion sogar eine dritte Monomer-Einheit (I-83mer) bilden können. Diese drei Monomer-Einheiten können miteinander im Gleichgewicht liegen, sodass die Intensitäten der einzelnen Banden je nach Gleichgewichtszustand von

I-83mer, 92- und 94mer mit den daraus möglichen Oligomeren stehen und keine kinetischen Rückschlüsse mehr zulässt.

- iv) Durch die reverse Transkription vom 92/94mer entsteht eine DNA, die über Größenabschätzung mittels DNA-Leiter die passende Länge ergibt.
- v) Bei Elektrophorese mit langem Migrationsweg trennen sich 92/94mer ansatzweise in zwei Banden.

Die darunter liegende RNA-Spezies ist das finale Spalt-Produkt, das lineare 83mer. Dies wird bestätigt durch die Eigenschaften der RNA Spezies:

- i) Beim Umschreiben in cDNA passt die Länge der cDNA zum Monomer zur DNA-Marke des DNA-Größenstandards (unter Berücksichtigung der Zunahme der Länge aufgrund der RT-PCR Primer).
- ii) Bei jeder Ribozym-Reaktion egal welcher RNA-Spezies aus dem System CRZ-2 bildet sich diese RNA immer.
- iii) Diese RNA organisiert sich ausschließlich in RNA-Spezies, die ohne Doppelbanden auftreten (im Gegensatz zur 92/94mer-RNA-Probe).
- iv) Über eine grobe Abschätzung mittels RNA Größenstandard befindet sich die Bande nicht weit unter der letzten RNA-Marke von 100 Basen, das 92/94mer liegt allerdings zwischen dem I-83mer und der 100er Marke.
- v) Diese RNA lässt sich nicht weiter spalten.
- vi) Diese RNA reagiert unter den Bedingungen der Ribozym-Reaktion ausschließlich zu längeren RNAs.

Gestützt werden die beiden Interpretationen durch die Ergebnisse der Analysen der Test-Sequenzen (s. Ergebnisse Teil II). PBD3 und 4 weisen terminale Segmente auf, die sich nach Abspaltung im Gegensatz zum CRZ-2-System, in der Länge nicht um zwei, sondern um sechs Basen unterscheiden, so dass die intermediären Spalt-Produkte im denaturierenden Polyacrylamid- (dPAA-) Gel getrennt und eindeutig zugeordnet werden können. Bei PBD3 und 4 treten die Spalt-Produkte I-83mer, 91mer und 97mer auf. Die oben als I-83mer interpretierte RNA zeigt ein identisches Laufverhalten zum I-83mer des Testsystems 4, das 91mer befindet sich minimal unter dem 92/94mer und das 97mer etwas oberhalb des Gemisches im Gel. Das parallele Auftragen der Test- und CRZ-2-RNAs in einem Gel ermöglichte das Abzählen der Basenzahlen vergleichbar zu einem RNA-Standard gemäß

der oben erfolgten Auslegungen zu den genannten möglichen Spalt-Produkten (s. Manuskript 3, Abb. S2d).

Die Analyse aller Ligationsprodukte basiert zunächst auf der Analyse der Ligationsprodukte des linearen 83mers, da die entsprechenden Multimere aus nur einer Monomereinheit aufgebaut sind (s. Manuskript 2, Abb. 4, Spalte 1 für eine typische Illustration des Ergebnisses einer Ribozym-Reaktion des I-83mers). Zunächst wurde zwischen intra- und intermolekularen Reaktionsprodukten unterschieden. Eine Analyse der einzelnen Reaktionsprodukte über Elution aus dem Gel, Ethanolpräzipitation und erneute Inkubation ergab, dass während der Isolationsprozedur erneute Spaltung und Ligation erfolgten. Das passierte sogar, wenn die Isolation der RNAs aus dem Gel in Wasser, alle experimentellen Schritte auf Eis und bei einer Umgebungstemperatur von 10 °C erfolgten (Daten nicht gezeigt). Daher wurde das gesamte Reaktionsgemisch auf denaturierenden Polyacrylamid-Gelen bei Raumtemperatur untersucht. Mit zunehmender RNA-Länge (z.B. nach Oligomerisierung) wird der Einfluss von Sequenzeffekten auf das Laufverhalten wahrscheinlicher, und es kann auch zur Bildung von komplexeren RNA-Strukturen kommen, sodass das Denaturierungs-Agens Harnstoff diese Strukturen nicht vollständig entfaltet. Dieser Umstand hat zur Folge, dass ein Abgleich mit einem RNA-Längen-Standard immer weniger verlässlich ist. Außer Polyacrylamid-Gelelektrophorese (PAGE) wurde auch Größenausschluss-Chromatographie und dynamische Lichtstreuung (DLS) zur Analyse herangezogen, was die Aussagekraft nicht verbesserte (Daten nicht gezeigt). Zwei Variationen der hier üblichen denaturierenden PAGE wurden durchgeführt, native PAGE und heiße denaturierende PAGE (55 °C), was die Aussagefähigkeit über die einzelnen RNAs des Produktgemisches ebenfalls nicht erhöhte (Daten nicht gezeigt). Zudem wurden die Migrationsverhaltensweisen der RNAs in dPAA-Gelen unterschiedlicher Konzentration, gemäß Ferguson gegenübergestellt (51). Für RNAs ist davon auszugehen, dass sich lineare Multimere in einem PAA-Gel zu ihrem Multimerisierungs-Grad proportional verhalten. Auffällig war, dass sich die Position der vermeintlich dritt längsten RNA im Gemisch relativ zu den anderen RNAs änderte und mit abnehmender dPAA-Konzentration sogar an die Position der zweit längsten gelang (Daten nicht publiziert). Es wurde angenommen, dass es sich bei dieser RNA um das zyklische 83mer handelte. Das Migrationsverhalten im dPAA-Gel von einer linearen und zyklischen RNA mit jeweils 83 Basen und hoher Sequenz-Ähnlichkeit wurde gegenübergestellt. Dafür wurde ein katalytisch inaktives Monomer entwickelt. Nach

enzymatischer Ligation ist nur ein zyklisches Produkt vorhanden, im Gegensatz zu der Vielzahl an Produkten nach der Ribozym-Reaktion des I-83mers. Die inaktive RNA war sequenzidentisch zum katalytisch aktiven I-83mer, außer, dass der für die katalytische Aktivität essentielle Loop A durch einen Doppelstrang ersetzt und am 5'-Terminus die Sequenz zu 5'-GGG AGA-3' variiert wurde. Zudem wurde die essentielle Base C25 durch ein Guanin ersetzt. Es zeigte sich im denaturierenden PAA-Gel, dass sich die zyklische Form über der linearen befindet (s. Manuskript 2, Abb. 4). Daher wurde angenommen, dass (mindestens) eine der RNAs, aus dem Produktspektrum der Ribozym-Reaktion, die über dem I-83mers liegen, zyklisch sein kann.

Der kanonische Weg zur Unterscheidung zwischen ringförmiger und linearer RNA besteht darin, eine exoribonukleolytische Hydrolyse der linearen RNA durchzuführen. Verwendet wurde die Phosphodiesterase I (PDE I) aus *Crotalus atrox*. Drei Wege wurden gewählt, um die RNA Topologie zu untersuchen: PDE I wurde auf ein RNA enthaltendes Gel gegeben, PDE I wurde in ein Gel sowohl in der ersten als auch zweiten Dimension einpolymerisiert und PDE I wurde in eine RNA enthaltende Lösung gegeben.

Die Verwendung der PDE I auf dem Gel und im Gel in der zweiten Dimension brachte keine neuen Erkenntnisse in Bezug darauf, welche der RNAs zyklisch sein könnte (Daten nicht gezeigt).

Übereinstimmend ergab die Analyse mittels PDE I in Lösung (s. Manuskript 2, Abb. 5) und im Gel (Daten nicht gezeigt), dass eine RNA-Spezies als Ring vorliegt. Ebenfalls zu beobachten war, dass die zyklische RNA auch hydrolysiert wurde. Dies ist für die Reaktion in Lösung einfach zu erklären, da die Ribozyme mit ihrer offenen bzw. geschlossenen Form im Gleichgewicht stehen. Somit wird den Gesetzmäßigkeiten von Le Chatelier folgend, die lineare aus den zyklischen und den Oligomeren-RNAs nachgebildet, sobald die lineare Form hydrolysiert wurde. Für einen Abbau der RNAs im Gel hingegen müsste innerhalb des Gels ausreichend Platz für die PDE I sein, RNA und Wasser aufzunehmen, sich umzufalten und ggf. zu reagieren. Ob ausreichend Platz vorhanden ist, wurde mit Molekulargewichts-Vergleichen nachgewiesen. Das Molekulargewicht der PDE I wurde mittels Abgleich mit Protein-Molekulargewichts-Standard bestimmt und beträgt ca. 50 kDa (Daten nicht gezeigt). Ein 50 kDa schweres im PAA-Gel befindliches Protein kann somit aktiv sein, obwohl dem Gel 7 M Harnstoff beigelegt wurde und für Hydrolyse-Reaktionen nur das Wasser aus dem Laufpuffer

zur Verfügung stand. Dagegen ist das lineare 83mer nur 25,6 kDa schwer. Schließt man in erster Näherung vom Molekulargewicht auf die Größe, hat die RNA ausreichend Platz um ihre Topologie zu ändern.

Die Tatsache, dass die vermeintliche zyklische RNA, wenn auch deutlich langsamer als das lineare Gegenstück, durch eine Exonuklease abgebaut wird, schwächt die Überzeugungskraft des erhaltenen Ergebnisses. Um eindeutig zu identifizieren, welche der RNAs des Produktspektrums zyklisch sind, wurde eine zwei-dimensionale (2D) PAGE angewendet. Es wurde erwartet, dass im Gegensatz zu den linearen RNAs, deren Migrationsverhalten im (d)PAA-Gel proportional zu ihrer Länge ist, das Laufverhalten der zirkulären RNAs von dieser Proportionalität abweicht (52,53). Ein Laufverhalten, das von dem eines linearen RNA Moleküls abweicht, muss nicht ausschließlich bedeuten, dass es sich um ein zyklisches RNA-Molekül handelt. Möglich ist grundsätzlich auch, dass verzweigte oder *genickte* RNAs (52) ein im Vergleich zu ihrer Länge atypisches Migrationsverhalten zeigen. Das Vorhandensein einer Verzweigung würde bedeuten, dass Ligationsreaktionen auch unter Nutzung der 2'-OH-Gruppe erfolgen. Dies ist auszuschließen aufgrund der inhärenten Hairpin-Ribozym-Aktivität. Die Ausbildung einer nicht vollständig denaturierten Sekundärstruktur der RNA im dPAA-Gel ist durchaus möglich und zeigt sich durch seltenes Auftreten einer Doppelbande des linearen 83mers in der 2D-PAGE. Da diese Spezies in hohem Maße auf kleinem Raum im Gel vorkommt, ist es wahrscheinlich, dass eine vollständige Denaturierung selbst bei Verwendung von 7 M Harnstoff nicht gewährleistet werden kann.

Da bei der nativen PAGE diverse RNAs gleiches Laufverhalten zeigen, wurde stets die denaturierende PAGE für die erste Dimension verwendet. Wenn in der zweiten Dimension auch denaturierende Bedingungen gewählt werden, wurde die Separation der RNAs aufgrund erhöhter PAA-Konzentration verstärkt. Zudem sollte eine RNA-Spezies mit bekannter Topologie mehrfach vertreten sein. Dies ist insbesondere dann nötig, wenn ein unbekanntes System untersucht wird. Im Falle einer gleichen Anzahl linearer und zyklischer Produkte wäre unklar, welches Set an RNAs die zyklischen darstellen, es entstünden zwei Diagonalen. Hier wurde sich für den Zusatz eines Größenstandards mit linearer RNA entschieden, der kommerziell erhältlich ist¹.

¹ Eine Anfrage per Email an den Anbieter ergab, dass die RNAs des Größenstandards ausnahmslos linear sind.

Die linearen RNA Spezies bilden in der zweiten Dimension eine Diagonale, wobei zirkuläre RNAs darüber oder darunter liegen. Mittels 2D-PAGE konnte zwischen den einzelnen RNAs im Produktgemisch nach Ligationsreaktion des I-83mers unterschieden werden (s. Manuskript 2, Abb. 6). Auch die Anwendung der nativen PAGE in der zweiten Dimension ermöglichte eine Differenzierung zwischen intra- und intermolekularen Ligationsprodukten. Die zirkuläre RNA migriert schneller als alle anderen Spezies durch das Gel, und ist dadurch einfach und eindeutig zu identifizieren. Zusätzlich ist hier die Unterscheidung zwischen *genickter* und völlig linearer bzw. kovalent geschlossener RNA möglich gewesen. Die native zweite Dimension zeigte eindeutig eine konstante Migrationsgeschwindigkeit des I-83mers unabhängig vom Denaturierungsgrad. Dagegen migrierte das zyklische 83mer wesentlich schneller durch das Gel und war auf der gleichen Höhe zu finden, wie das I-83mer (s. Manuskript 3, Abb. S3b). Somit konnte eine zyklische RNA, das zyklische 83mer, nach Ribozym-Reaktion des finalen Spalt-Produktes gebildet und eindeutig identifiziert werden.

3.1.2 Analyse der Multimere

Die Methode der 2D-Gel-Analyse, sowohl mit nativer als auch denaturierender zweiter Dimension wurde bei dem Produkt-Spektrum der Ribozym-Reaktion von CRZ-2 angewendet. Im Gegensatz zur oben beschriebenen Situation, in der das entsprechende isolierte l-83mer einer Ribozym-Reaktion unterworfen und das zyklische Monomer gefunden wurde, konnte für CRZ-2 selbst keine zyklische RNA nachgewiesen werden (s. Manuskript 3, Abb. S3c, d). Ein gemeinsames Auftragen von CRZ-2 und dem linearen 83mer in einem Gel zeigt, dass sich keine RNA auf der Höhe des zyklischen 83mers in der Spur des CRZ-2 befindet (s. Manuskript 3, Abb. 4).

Daher war es möglich, die übrigen linearen RNAs initial über den Vergleich mit einem linearen RNA Größenstandard der Länge nach zuzuordnen. Da dies aber kein hinreichender Beweis für eine tatsächlich erfolgte Multimerisierung war, wurde eine Sequenzierung der zu den RNAs komplementären DNA (cDNA) durchgeführt. Im folgenden Abschnitt (bis S. 16 einschließlich des zweiten Absatzes) werden Experimente und Ergebnisse aus Manuskript 2 diskutiert und zusammengefasst.

Für die reverse Transkription (RT) wurden die RNAs aus der Ribozym-Reaktion verwendet, die durch langsames Migrationsverhalten und Ausschluss der zyklischen Struktur im dPAA-Gel als RNA-Oligomer angenommen wurden. Diese RNAs waren gemäß Abschätzung mit RNA-Größenstandard ca. 260 bis 320 Basen lang. Nach RT-PCR ergaben sich allerdings DNAs einer Länge von ca. 500 bp im nativen PAA-Gel, was nicht durch die Verlängerung der Sequenz mittels RT-PCR-Primern (je 9 Nukleotide pro Terminus) zu erklären war. Die Analyse der gleichen DNA im denaturierenden Gel ergab einen „dünnen Schmier“, der sich sogar weit über die 500 bp-Markierung hinauszog. Der DNA-Nachweis mit Längen von 300 bp bis zum oberen Gel-Rand erinnerte an Analysen alkalischer Hydrolysen von RNAs und an *in vitro* Transkriptionen. In beiden Fällen erscheint RNA als Schmier im dPAA-Gel, einmal als statistische Abbauprodukte und einmal als unvollständige Syntheseprodukte, wobei augenscheinlich alle RNA-Fragmente von einem Nukleotid bis zur maximalen Anzahl Nukleotide des Zielmoleküls zu finden sind. Da die Bedingungen der reversen Transkription und der Ribozym-Reaktion ähnlich sind (Anwesenheit von Mg²⁺, 37 °C, und eine Reaktionszeit von mindestens 30 Min.) ist es möglich, dass sich einige RNA-Moleküle zum Ring prozessiert haben und durch die reverse Transkriptase permanent abgelesen wurden.

Dies ist möglich, weil die verwendete reverse Transkriptase ursprünglich einem Virus zur DNA-Synthese während der *rolling-circle* Replikation diente, und auch dort RNA-Ringe revers transkribierte. *In vitro* wird die Reaktion mit Hitze-Inaktivierung der Transkriptase beendet. Es wird zufällig der Endpunkt der Synthese bestimmt, sodass die resultierenden cDNAs unterschiedlicher Längen im dPAA-Gel als Schmier wiederzufinden sind.

Eine als Nebenreaktion angenommene unbeabsichtigte Selbst-Prozessierung der RNA-Substrate im Ansatz der reversen Transkription erklärt die deutlich verlängerten distinkten Banden im nativen PAA-Gel. Die RNAs haben die Möglichkeit sich selbst zu spalten und sich intra- oder intermolekular zu ligieren. Vermutlich wurden intermolekulare längere cDNAs erhalten und diese amplifiziert. Um dies zu testen, wurde ein Aliquot der ursprünglich isolierten RNAs einer Ribozym-Reaktion unterworfen und einer reversen Transkription mit angeschlossener Hydrolyse der entstandenen DNA in einem dPAA-Gel gegenübergestellt. Die entstandenen RNA-Spezies beider Experimente zeigten ein identisches Migrationsverhalten im PAA-Gel (Daten nicht gezeigt), was die Annahme stützt, dass während der reversen Transkription Ribozym-Reaktionen ablaufen. Somit ist es unwahrscheinlich, dass die cDNAs den RNAs entsprechen, die aus dem dPAA-Gel initial isoliert wurden. Dennoch sind sie Produkte der Ribozym-Reaktionen des zu untersuchenden Systems und wurden für die Sequenzierung eingesetzt.

Es wurden 47 Sequenzen analysiert und 30 Monomere, 16 Dimere und 1 Trimer herausgelesen (s. Manuskript 2 *Gene bank indices: EMBL Nucleotide Sequence Database*, Trimer: Zugangsnummer: HF565114; exemplarisches Dimer 1: Zugangsnummer: HF565115; Exemplarisches Dimer 2: Zugangsnummer: HF565116).

Ein Großteil der DNAs wies Mutationen, wie Deletionen und Insertionen auf. Das Auftreten war unabhängig vom eingesetzten Sequenzierprimer, Oligomerisierungsgrad der untersuchten Sequenz oder der Anordnung der Segmente innerhalb eines Oligomers. Eine fehlerhafte Synthese durch die reverse Transkriptase ist unwahrscheinlich, da sie bei 37 bis 55 °C nur eine Fehlerrate von ca. $3,6 \text{ bis } 1,8 \times 10^{-5}$ aufweist. Auch bei der angeschlossenen PCR ist die Fehlerrate durch Verwendung der Vent_R [®] DNA proof-reading Polymerase minimal (Mutationsfrequenz aller Klassen von DNA Polymerase Fehlertypen: Transitionen, Transversionen und Leserasterverschiebungen ca. 25×10^{-4})(54). Es gab jedoch zwei Amplifikationsereignisse, auf die kein besonderer Einfluss ausgeübt werden konnte: die

Amplifikation des inserttragenden Plasmids innerhalb des *E. coli* nach der Transformation und die Syntheseschritte bei der Sequenzierung durch die Firma GATC Biotech AG. Tatsächlich birgt jede Grundeinheit eines Multimers (die Sequenz des I-83mers) einen Mutations-*Hot-Spot*. Das sind Sequenzbereiche, die aus mehrfachen Basen-Wiederholungen bestehen. Diese Wiederholungen sind für Polymerasen schwierig zu quantifizieren, d.h. statt vier Adeninen werden ggf. drei oder fünf erkannt und somit mehr oder weniger Basen in den Tochterstrang eingebaut. Da sich im 83mer sieben dieser *Hot-Spots* befinden, war das Auftreten von Mutationen sehr wahrscheinlich. Für ein etwaiges rationales *Design* neuer Ribozyme, wobei letztendlich eine Sequenzierung zur Multimeren-Analyse angeschlossen werden soll, sollte die Vermeidung solcher *Hot-Spots* berücksichtigt werden.

Monomere, die aus Mehrfach-Insertionen hervorgehen, waren durch Wiederfinden der Primer-Sequenzen leicht von den realen Selbst-Prozessierungs-Ergebnissen zu unterscheiden. Ist das potenzielle Trimer nur von Primer-Sequenzen flankiert, handelt es sich um ein reales Trimer. Ist die Sequenz hingegen von Primerteilen unterbrochen, handelt es sich um wiederholte Insertionen eines Monomers, Dimers etc.

Da ausschließlich DNA mit Längen von 250 bis 500 bp aus dem Gel isoliert wurden, war überraschend, dass überhaupt Monomere vertreten waren. Es ist vorstellbar, dass die Trimere nicht mit weiteren Trimeren einen Doppelstrang bilden, sondern auch mit einem Dimer und einem Monomer oder sogar mit drei Monomeren. Da die Monomere, insbesondere das I-83mer, in jeder Ribozym-Reaktion im Überschuss vertreten waren, ist die gemischte Zusammensetzung eines Doppelstranges besonders wahrscheinlich und spiegelt sich im Segmentverhältnis wider.

Die Monomere, Dimere und Trimere bestanden aus 83- und 92meren, wobei Oligomere, die am 5'-Ende ein 94mer aufweisen, nicht erfasst wurden. Dies war beabsichtigt, weil ausschließlich ein Primer-Paar verwendet wurde, das das I-83mer erkennt und nebenbei das 92mer aufgrund der identischen Termini (s. Manuskript 2, Material und Methoden). Für eine vollständige Analyse der Oligomere hätten Primer entwickelt werden müssen, die den Einbau aller möglichen Monomere (83-, 92- und 94mer) in allen Positionen innerhalb eines Oligomers abdecken. Die Verwendung eines Primer-Paares für das 83mer erschien besonders sinnvoll, da alle Oligomere diese Einheiten aufweisen und so die DNA-Ausbeuten maximal sind, unerheblich aus welchen Monomeren die Oligomere aufgebaut sind. Da die

Mengen an langen DNAs derart gering ausfielen, dass eine Vielzahl von PCRs nötig war um Klonierungsversuche anzuschließen, wurde nicht versucht spezifischere Primer-Paare zu verwenden. Ein interner Einbau des 94mers in ein Oligomer ist durch das am 5'-Ende befindliche Triphosphat nicht möglich und wurde demzufolge nicht gefunden. Da die anfängliche Oligomerisierung über gefundene Dimere und Trimere demonstriert war, wurde die Nachweis-Erbringung mittels Sequenzierung beendet.

Intuitiv würde man auf die RNA-Sequenzierung zurückgreifen, um die oben erwähnten Schwierigkeiten zu umgehen. Dies ist hier allerdings nicht anwendbar, da insbesondere längere Sequenzen (wie Oligomere) für den Sequenzier-Prozess zerstückelt, amplifiziert und durch Überlappungsbereiche wieder zu einer Sequenz rekonstruiert werden würden. Diese Technik ist bei der Multimeren-Analyse nicht zweckdienlich, da jede Monomer-Einheit einen potenziellen Überlappungsbereich aufweist und die exakte Multimer-Rekonstruktion höchst wahrscheinlich misslingen würde. Die DNA-Sequenzierung dagegen ist nachteilig, da zeitaufwendig. Zudem finden Spalt- und Ligations-Reaktionen auch während der reversen Transkription statt, was die Ausbeute der zu untersuchenden RNAs stets reduziert und es ist so nur ein indirekter Nachweis möglich. Da während der RNA-Isolation und reversen Transkription bereits Selbst-Prozessierung stattfindet, steht sehr wahrscheinlich nicht die ursprüngliche RNA resultierend aus der beabsichtigten Ribozym-Reaktion für die weiteren Schritte bis zur Sequenzierung zur Verfügung. Auf der Suche nach einer geeigneten Alternative war die direkte Visualisierung des Reaktionsgemisches mittels Einzelmolekül-Analyse über *atomic force microscopy* (AFM) die Methode der Wahl.

Erste AFM-Aufnahmen von Hairpin-Ribozym-Varianten wurden 2001 von J.M. Burke publiziert (55). Die hier erhaltenen Aufnahmen wurden in Kooperation mit der Arbeitsgruppe von Dr. Delcea durch Dr. Block erstellt (Universität Greifswald). Beide Referenz-Systeme (CRZ-2 und das I-83mer) wurden einzeln nach Ribozym-Reaktion abgebildet. Es wurden Monomere, Dimere und Trimere gefunden (s. Manuskript 3, Abb. 7, 8a und S4a, b, S5a, b).

Bei der Analyse der 83 Basen langen Monomere konnte nicht zwischen zyklischer und linearer RNA unterschieden werden. Das hätte bedeutet, dass es möglich ist, mittels AFM zwischen der An- und Abwesenheit einer einzelnen kovalenten Verknüpfung zu unterscheiden, was nicht der Fall war.

3.2 Fazit Ergebnis-Teil I

Für CRZ-2 konnte durch Anwendung der 2D-PAGE, Sequenzierung und AFM direkt und indirekt nachgewiesen werden, dass die Spalt-Kaskade vollständig abläuft. Anfängliche Oligomerisierungen sind nachweisbar, aber kein mit Ethidium-Bromid im dPAA-Gel, oder per AFM detektierbares zyklisches 83mer wurde gefunden.

Für die Analyse des isolierten I-83mers wurden folgende Methoden kombiniert:

- i) 2D-PAGE (sowohl in beiden Dimensionen denaturierend und unter Verwendung eines Standards für lineare RNA, als auch nur in der ersten Dimension denaturierend und in der zweiten Dimension nativ)
- ii) exonukleolytischer Abbau von RNA in Lösung und im Gel
- iii) Vergleich des Laufverhaltens eines inaktiven RNA-Monomers mit dem Reaktionsgemisch des I-83mer nach Ligationsreaktion
- iv) AFM
- v) Ferguson-Plot.

Damit war es möglich zu bestimmen, dass bei der Ribozym-Reaktion des linearen 83mers unter Hairpin-Ribozym-Standard-Bedingungen eine zyklische RNA, das 83mer, und lineare Monomere, Dimere und Trimere entstanden sind.

Somit stand für die weitere Analyse von anderen Hairpin-Ribozym-Varianten eine Vielzahl von Methoden zur Verfügung, um Selbst-Prozessierungen zu untersuchen.

3.3 Teil II: Ergebnisse der bioinformatisch entworfenen Test-Sequenzen

3.3.1 Bioinformatischer Entwurf von Hairpin-Ribozym-basierten Test-Sequenzen

Beim bioinformatischen Entwurf durch die Kooperationspartner aus der Arbeitsgruppe von Prof. Ivo Hofacker wurde das *Vienna RNA package* 2.0 benutzt (47). Dabei werden Strukturen der minimalen freien Energie und deren Zustandssummen berechnet. Mit Hilfe der Zustandssummen wurden für jeden Schritt in der Reaktionskaskade Wahrscheinlichkeiten für die Ausbildung reaktiver Sekundärstrukturen berechnet (s. Manuskript 3, Gleichung 1). Der geschwindigkeitsbestimmende Schritt der Kaskade ist der Schritt mit der geringsten Wahrscheinlichkeit zur Ausbildung einer katalytisch aktiven Sekundärstruktur. Dafür müssen beide möglichen Wege (entweder wird zuerst der 3'-, oder der 5'-Terminus abgespalten) berücksichtigt werden.

Es sollten Ribozyme entwickelt werden, die a) die Ausbeute an zyklischem Monomer und b) die Ausbeute an zyklischem Dimer maximieren. Für das erste Szenario wurden die Minima aus beiden Wahrscheinlichkeiten für die Reaktionskaskade mit dem Quotienten der Zustandsfunktion für die Zyklisierung berechnet und mit dem Anteil an zyklischen Molekülen im Gleichgewicht multipliziert (s. Manuskript 3, Gleichung 2). Das Programm *switch.pl* (46) aus dem *Vienna RNA package* 2.0 (47) wurde verwendet um bistabile RNAs zu entwerfen. Dabei wird berücksichtigt, dass die für die Katalyse relevanten Basen in den Loops nicht ausgetauscht werden. Alle anderen Positionen eines etwaigen neuen Ribozyms wurden durch *Switch.pl* mutiert, also theoretisch neue Basen angenommen und damit 10.000 Sequenzen entworfen und anschließend mit den oben beschriebenen Gleichungen evaluiert.

Ebenso wurde für die Entwicklung von Test-Ribozymen vorgegangen, die vornehmlich zyklische Dimere nach der Spalt-Kaskade hervorbringen sollen. Dabei wurde berücksichtigt, dass sich auch *full-length* Ribozyme erst (nicht-kovalent) dimerisieren können und dann die Spalt-Kaskade durchlaufen wird. Wieder wurden Zustandssummen gebildet und daraus die Wahrscheinlichkeiten aktiver Sekundärstrukturen berechnet (s. Manuskript 3, Gleichung 3). Die Berechnung der Bildung von zyklischen Dimeren unter Berücksichtigung einer Konzentration im unteren nanomolaren Bereich erfolgte nach Bernhart (56) gemeinsam mit den Minima der Dimer-Reaktionskaskade in der zweiten Optimierungsfunktion (κ_2) (s. Manuskript 3, Gleichung 4). Die Ergebnisse der Funktionen sind dimensionslose Zahlen,

wobei gilt, je höher der Wert, desto wahrscheinlicher ist die gesuchte Eigenschaft beim Test-Ribozym vertreten (s. Manuskript 3, Tabelle 1).

Vier Systeme unterschiedlicher Eigenschaften wurden für die experimentelle Analyse ausgewählt.

Tabelle 1: Eigenschaften der Test-Systeme und CRZ-2 gemäß bioinformatischer Charakterisierung

RNA	Länge	Fragment Länge	Zyklisierungstendenz	
			Monomere	Dimere
CRZ-2	103	11 + 83 + 9	sehr gering	sehr gering
PBD1	103	11 + 83 + 9	sehr hoch	sehr hoch
PBD2	103	11 + 83 + 9	sehr hoch	gering
PBD3	105	8 + 83 + 14	moderat	moderat
PBD4	105	8 + 83 + 14	moderat	gering

Die CRZ-2-Synthese-Strategien vom Klenow-Primer bis zur selbst-prozessierten RNA wurden für die Test-Sequenzen beibehalten. Obwohl die Sequenz-Identitäten eher moderat sind, von 53,4 % bis 58,3 % (Manuskript 3, Abb. 3b) waren alle 103/105mer an der gleichen Stelle im Gel zu finden (Daten nicht gezeigt). Analog zum Referenz-System CRZ-2, kam es auch bei den Test-Sequenzen bereits während der Synthese über *in vitro* Transkription zur Selbst-Prozessierung der RNAs.

3.3.2. Vergleich der vorhergesagten Eigenschaften mit den experimentellen Ergebnissen

3.3.2.1 Übereinstimmungen

Bei allen Test-Sequenzen konnte das zugehörige zyklische 83mer über 2D-PAGE identifiziert werden (Manuskript 3, Abb. 5). Somit wurde das erfolgreiche Durchlaufen der Spalt-Kaskade und die Monomer-Zyklisierung aller Test-Ribozyme indirekt nachgewiesen. Wie im ersten Ergebnis-Teil erwähnt, konnten alle Zwischenstufen der Spalt-Kaskade über die Test-Sequenz 4 im dPAA-Gel gezeigt werden (s. Manuskript 3, Abb. S2d). Über die Referenz-Systeme und PBD4 war es auch möglich, die Spalt-Produkte der anderen Test-Systeme zuzuordnen und so den vollständigen Ablauf der Kaskade direkt zu bestätigen.

Alle Test-Systeme sollten im Gegensatz zu CRZ-2 nach Ribozym-Reaktion zyklische Monomere ausbilden können. Unter der Annahme, dass im entstandenen Schmier der 2D-Analyse von PBD4 auch zyklische Monomere vorhanden sind (s. S. 21 für eine Erklärung

dieser Interpretation), ist auch diese Vorhersage korrekt. Die Zuordnung der zyklischen RNAs zum Gel der ersten Dimension wurde benutzt, um mit dem Programm *Gene ImagIR V. 4.05* Aussagen über die Menge der einzelnen zirkulären RNA Spezies nach Ribozym-Reaktion zu machen. Mit den Anteilen an zyklischer RNA von PBD1: 43 %, PBD2: 50 %, PBD3: 8 % und PBD4: 2 % stellte sich heraus, dass die vorhergesagte Tendenz zur Monomer-Zyklisierung korrekt war. Die einzelnen RNA-Spezies von PBD4 waren schwierig zu quantifizieren. In der entsprechenden 2D-PAGE war anstatt eines eindeutigen *Spots* für die zyklische RNA ein ausgedehnter Schmier überall gleicher, schwacher Intensität im dPAA-Gel zu finden. Ein Erklärungsvorschlag für das Zustandekommen des Schmiers ist auf S. 21.

Für PBD1 wurde angenommen, dass es das aktivste Ribozym der vier Test-Systeme darstellt, was durch das Fehlen sämtlicher Spalt-Produkte deutlich wird (s. Manuskript 3, Abb. 4). Die entstandenen Ligationsprodukte stellen wahrscheinlich ein lineares Dimer (vorzugsweise 83+83 Basen), und nicht näher bestimmbar ein weiteres lineares Dimer oder lineares Trimer (vorzugsweise 83+83+83 Basen) dar. Die AFM-Analyse konnte diese mögliche Interpretation bestätigen, aber keinen finalen Beweis erbringen. Ein Dimer bzw. Trimer, welches ausschließlich ein anderes Spalt-Produkt als das I-83mer aufweist, ist neun Basen entsprechend 1,8 nm länger. Dieser Abstand ist zu klein, um mittels AFM zwischen den unterschiedlichen möglichen Dimeren und Trimeren differenzieren zu können. Die Auflösungsgrenze ist von der Breite des Cantilevers abhängig, der die RNA-Proben ertastet. Es wurden ausschließlich Cantilever verwendet, deren Breite kleiner als 5 nm waren und durch Tests an Kalibrier-Proben durch Dr. Block vorab getestet (s. Manuskript 3, Diskussion). Es war möglich lineare Trimer als längste lineare Oligomere zu identifizieren (s. Manuskript 3, Tabelle 3 und Abb. S5c).

Eine Voraussetzung für eine Zyklisierung eines Dimers ist das einmalige Ligieren zweier Monomere zum offenen Dimer. Die in Tabelle 1 angegebenen Werte können somit in erster Näherung auch für eine Vorhersage zur Tendenz einmalig kovalent-verknüpfter Dimere herangezogen werden.

Die Fähigkeit von PBD2, 3 und 4 zu dimerisieren soll geringer ausgeprägt sein als bei PBD1. Vergleicht man die Intensitäten der entsprechenden Banden (s. Manuskript 3, Abb. 4), ist diese Vorhersage augenscheinlich korrekt.

3.3.2.2. Unterschiede

Um abzuschätzen, wo sich ein zyklisches Dimer (aus 83+83mer) im dPAA-Gel befinden würde, wurde eines auf der Sequenz-Basis von CRZ-2 bioinformatisch entworfen. Es enthielt eine ähnliche Basen-Komposition, um ein zum realen CRZ-2-Dimer möglichst ähnliches Laufverhalten im Gel zu ermöglichen. Die für die Katalyse essentiellen Loops wurden verändert (s. Manuskript 3, Material und Methoden). Somit war das Referenz-Dimer nicht mehr selbst-prozessierend und musste enzymatisch ligiert werden, was ausschließlich ein Produkt ergab. Dieses war nicht mittels Exonuklease hydrolysierbar (Manuskript 3, Abb. 6) und wurde daher als zyklisches Dimer interpretiert. Es befand sich auf der Höhe von 800 Basen des RNA-Größenstandards. Dort wurde bei keinem Test-System eine RNA gefunden. Da auch die Dimere aller Systeme zyklisieren sollten, aber jeweils nur ein deutlicher Spot auf der Höhe des zyklischen Monomers gefunden wurde, haben gemäß 2D-Analyse alle Test-Systeme gemeinsam, dass dieser Aspekt mittels wahrscheinlichkeitsbasiertem Entwurf nicht ausreichend genau beschrieben werden konnte.

Bei PBD2 könnte ein zirkuläres Trimer entstanden sein, dass bei der 2D-PAGE als *Spot* mit geringer Intensität zu finden war (Manuskript 3, Abb. 5). Ein zyklisches Dimer konnte ausgeschlossen werden, da die Bande nicht auf einer Höhe von 800 Basen gefunden wurde. Das vermeintliche Trimer ist somit im Gegensatz zum Dimer groß genug, um über eine Vielzahl von Freiheitgraden Verknäulungen zu ermöglichen, die ein schnelleres Laufverhalten zur Folge haben, als es das zyklische Dimer aufweist. Möglich wäre auch, dass es sich um eine *genickte*, nicht vollständig denaturierte, aber dennoch lineare RNA handelt. Der Annahme widerspricht die Tatsache, dass die Menge an umgebender RNA nicht besonders groß ist, sodass eine lokale ungenügende Denaturierung, die zu einer RNA-Struktur mit *nick* führt, unwahrscheinlich ist.

Durch die Einzel-Molekül-Analyse mittels AFM konnte eine Erklärung für das Vorhandensein des Schmiers in der 2D-Analyse bei PBD4 vorgeschlagen werden. Zusätzlich zu einer großen Anzahl von Monomeren und Dimeren die durch AFM illustriert werden konnten, wurden diverse RNA-Ringe unterschiedlicher Größe und Struktur dargestellt. Diese Ringe sind so groß, dass Verknäulungen als Ursache für das schnelle Migrieren durch das dPAA-Gel angenommen wurden, sodass die RNAs unterhalb der 800 Basen-Marke gefunden wurden. Die Tendenz zur Dimerzyklisierung sollte zu der von CRZ-2 vergleichbar gering sein. Mittels

AFM wurden jedoch RNA-Ringe gezeigt, deren Längen über den Dimerisierungsgrad hinausgehen (s. Manuskript 3, Abb. 8c). Somit konnte diese bioinformatische Charakterisierung nicht experimentell bestätigt werden. Es sei an dieser Stelle noch einmal darauf hingewiesen, dass mit AFM nicht das Vorhandensein oder Fehlen einer kovalenten Bindung gezeigt werden kann. Es besteht somit die Möglichkeit, dass die Enden der vermeintlichen Ringe ohne kovalente Verknüpfung zufällig beieinander liegen.

Grundsätzlich sollte die Dimerbildung bei CRZ-2 weniger stark ausgeprägt sein als bei den Test-Systemen (weiterhin unter der Berücksichtigung, dass die Tendenz zur Dimerzyklisierung auch in erster Näherung für die Bildung offenkettiger Dimere gilt). Es sind aber eine Vielzahl von linearen Oligomeren erkennbar, die alle im Gel-Abschnitt liegen, wo Dimere zu finden sind und unter denen auch eine Bande mit erhöhter Intensität ist. Somit ist die katalytische Aktivität der Referenz nur zum Teil bioinformatisch korrekt charakterisiert. Die Dimerzyklisierung blieb bei CRZ-2 wie vorhergesagt aus. Die Dimerisierung an sich verlief in bioinformatisch unerwartet hohem Ausmaß.

3.4 Struktur-Funktionsbeziehungen und Ausblick

Eine erste Diskussion bezüglich der berechneten Energiewerte der einzelnen Schritte der Spalt-Kaskade und den resultierenden Wahrscheinlichkeiten ist im Manuskript 3 zu finden und wird Teil der Dissertation von Stefan Badelt sein. In der vorliegenden Arbeit werden mögliche biochemische Ursachen für die auftretenden Ribozym-Aktivitäten untersucht.

Die vorgegebene Sequenz-Identität durch Konservierung der essentiellen Basen beträgt 37 %. Die Primärstrukturen und das Sequenz-Alignment sind in Manuskript 3, Abb. 3b gezeigt. Die Sequenz-Identitäten wurden für jede einzelne Test-Sequenz zu CRZ-2 bestimmt und liegen bei 53 bis 58 %. Da dieser Bereich über dem der Vorgabe durch Beibehaltung der essentiellen Loop-Sequenzen liegt, wurde kontrolliert, wie hoch die Sequenz-Identität aller Systeme ist. Ein Alignment aller Test-Sequenzen und CRZ-2 ergab eine tatsächliche Übereinstimmung von 41 %, was zusätzlichen vier Basen entspricht. Dies wird im Schema der Abb. 3b in Manuskript 3 deutlich.

Da die vier Basen nach dem wahrscheinlichkeitsbasierten Entwurf in allen Test-Sequenzen auftreten, kann man davon ausgehen, dass die vier Basen nicht zufällig gewählt, sondern wichtig für die Funktionalität der Systeme sind. Wird von den gemeinsamen Ähnlichkeiten in

der Primärsequenz auf die Ähnlichkeiten der Funktionen geschlossen, könnten die Basen essentiell für die Spalt-Aktivität, die Monomer-Zyklisierung, oder grundsätzlich für das Zustandekommen der notwendigen Helices sein. Diese vier Positionen sind exemplarisch für CRZ-2 in rot in Abb. 2 markiert².

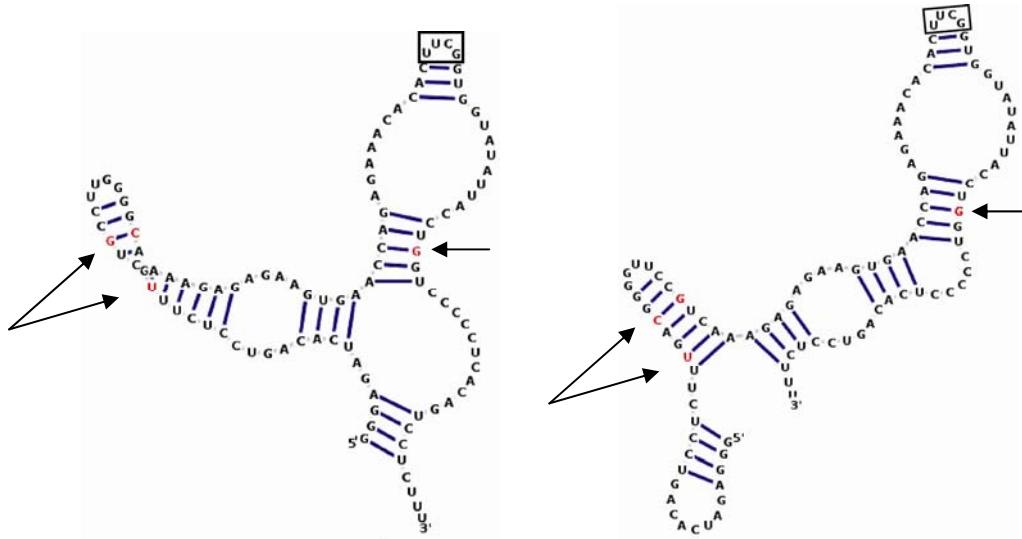


Abb. 2: Positionen der vier zusätzlichen Basen in den Sekundärstrukturen der beiden Konformationen von CRZ-2 gezeigt und in rot markiert; links: Sekundärstruktur zur Spaltung des 5'-Terminus, rechts: Sekundärstruktur zur Abspaltung des 3'-Terminus. Die Position des superstabilen Tetraloops ist eingerahmt.

Es handelt sich in allen Strukturen um ein GC-Basenpaar [23/32], ein Guanosin an der Position 82 und Uridin an Position 20. Das G82 und das U20 sind bei CRZ-2 immer in einer Helix über Watson-Crick- oder *Wobble*-Basenpaarung lokalisiert.

Das GC-Basenpaar ist ebenfalls in den 83meren vorhanden. Es könnte somit eine essentielle Rolle bei den Ligationen (intra- und oder intermolekular) spielen. Denkbar wäre wie oben erwähnt, dass das Basenpaar für die Ausbildung einer Helix, die als strukturgebendes Element dient, kritisch ist. Ein Watson-Crick-GC-Basenpaar weist drei Wasserstoffbrücken auf. Es wäre zukünftig interessant zu untersuchen, wie sich die Funktion des Ribozyms ändert, wenn die Interaktion an dieser Stelle geschwächt wird. Dies könnte z. B. durch Doppel-Substitution zu einem A-U-Basenpaar oder einem *Wobble*-Basenpaar erreicht werden (zwei Wasserstoffbrücken). Ob die Position wichtig ist, um grundsätzlich die Helix auszubilden, könnte durch Einführung einer Basen-Ausstülpung getestet werden. Dabei wäre

² Sekundärstruktur-Vorschläge bereitgestellt von Stefan Badelt (Universität Wien) unter Verwendung des Programms jViz (57).

interessant, wie sich diese Mutationen auf die Ergebnisse der Optimierungsfunktionen auswirken und ob sich im Experiment die Veränderungen erkennen lassen. Das Methoden-Repertoire könnte zur Untersuchung der Helix-Stabilität um UV-Schmelzkurven-Analysen erweitert werden. Wird die Anwesenheit des Basenpaars als wichtig bestätigt, sind diese Erkenntnisse für zukünftiges rationales *Design* neuer Hairpin-Ribozym-Varianten wertvoll.

CRZ-2 und PBD4 unterscheiden sich drastisch beim Zyklyserungsverhalten, was anhand der Ergebnisse der Optimierungsfunktionen nicht erwartet wurde. Es ist wahrscheinlich, dass die Sequenz des superstabilen Tetraloops maßgeblich ist. Dabei gilt, je mehr Basenübereinstimmungen im Tetraloop bei den Test-Systemen zu CRZ-2 vorhanden sind, desto ähnlicher ist das Verhalten zu CRZ-2. Entsprechend sind im Vergleich zu CRZ-2 alle Test-Ribozyme aktiver und der Tetraloop ist massiv verändert. Bei PBD2 ist er völlig substituiert und bei PBD1, 3 und 4 nur eine der vier Basen vertreten. Jedes Test-System hat somit seinen eigenen Tetraloop, der nicht superstabil sein muss.

Bei den Alignments der Test-Sequenzen untereinander ist eine Besonderheit aufgefallen. PBD3 und 4 unterscheiden sich nur in zwei Basen, sie haben aber entgegengesetzte Eigenschaften bezüglich der Dimer- bzw. Multimer-Zyklyse. Es handelt sich um die Base 65, das ist Base 1 des superstabilen Tetraloops und ist ein Uridin in CRZ-2 und PBD3. Bei PBD4 dagegen handelt es sich um ein Guanin. Bei CRZ-2 wurden keine RNA-Ringe gefunden und die Menge an Ringen bei PBD3 war sehr gering, wobei es sich sehr wahrscheinlich um das zyklische Monomer handelte. Ob es bei PBD3 auch zur Multimeren-Ringbildung gekommen ist, konnte nicht mit AFM untersucht werden. Falls Ringe mit AFM eindeutig identifizierbar wären, so wäre das Ausmaß deutlich reduziert im Vergleich zu PBD4, denn bei PBD4 konnten sie bereits im Gel detektiert werden. Somit scheint die erste Position des Tetraloops essentiell für die Multimeren-Bildung und Zyklyse zu sein. Interessant wäre es zu klären, welchen Einfluss der Einbau eines Cytidins an dieser Position hat. Ein Cytidin ist bei keinem der Ribozyme an der Stelle im Loop zu finden. Fraglich ist auch, warum gerade die erste Position so wichtig ist. Da es sich um eine Base in einem Einzelstrang-Bereich handelt, liegt eine tertiäre Interaktion als Erklärung nahe. Dies konnte beim bioinformatischen Entwurf nicht berücksichtigt werden und daher könnte das unerwartete Verhalten des Test-Systems PBD4 erklären. Gemäß Sekundärstruktur-Vorhersage hätte PBD4 kaum Dimere zyklysern sollen, übertrifft experimentell aber das erwartete Maß. Ob es sich

tatsächlich um eine tertiäre Interaktion handelt, könnte über Aufklärung der Tertiärstruktur ermittelt werden. Da die selbst-prozessierenden Systeme insbesondere in Lösung hoch aktiv und strukturell durchaus flexibel sind, kommen NMR, Röntgenstruktur-Analyse und AFM nicht ohne Weiteres in Frage. Es wäre jedoch möglich über *cross-links* die vermutete Interaktion, ausgehend vom Guanosin, zu fixieren und dann mit NMR zu verifizieren.

Der zweite Unterschied zwischen PBD3 und PBD4 betrifft die Base 3 des Tetraloops, somit Position 67 bei CRZ-2. Hier liegt bei CRZ-2 und PBD4 ein Cytidin vor, während PBD3 ein Uridin aufweist. Dieser Austausch im superstabilen Tetraloop bei Hairpin-Strukturen ist bereits in der Literatur beschrieben worden (29). Er bewirkt eine Reduktion der Hairpin-Schmelztemperatur, einen Anstieg der Gibbs'schen freien Enthalpie bei der Bildung des Hairpins und somit eine Destabilisierung der Sekundärstruktur bei 37 °C. Beim 5'-UUCG-3' Tetraloop kommt es zur Interaktion der 2-Amino-Funktion am Cytidin mit dem Phosphat zwischen den beiden Loop-Uridinen, wobei sich Cytidin in anti-Stellung befindet. Wird das Cytidin zum Uridin substituiert, müsste die Carbonyl-Gruppe mit dem Phosphat wechselwirken, was die erwähnte Destabilisierung bewirkt (58). Ob das bei den hier verwendeten Loops auch der Fall ist, könnte mit NMR-Untersuchungen (der Einfachheit halber nur mit Loops und kleinem angrenzenden Helix-Bereich) geklärt werden.

Kanonisch wäre es anzunehmen, dass die Monomer-Destabilisierung eine Öffnung der Struktur zur Folge hat und sich ein weiteres Molekül anlagern kann. Das müsste sich in einer Vielzahl von Dimeren und Oligomeren widerspiegeln, PBD4 weist aber augenscheinlich mehr Dimere als PBD3 auf (Manuskript 3, Abb. 4, vgl. Gelspuren 5 und 6). Die Destabilisierung der Sekundärstruktur hat somit eine Aktivitäts-Einschränkung zur Folge, so dass es weniger häufig zur Oligomerisierung kommt. Die Annahme, dass es sich tatsächlich um einen Effekt auf der Ebene der Sekundärstruktur handelt, wird dadurch gestützt, dass in diesem Fall die bioinformatische Charakterisierung stimmt.

Zukünftig ist angedacht, die Selbst-Prozessierung auch kinetisch zu erfassen. Dies bedeutet zunächst auf bioinformatischer Ebene die Analyse von Faltungs dynamiken und auf experimenteller Ebene das Finden eines geeigneten Inhibitors der Prozessierung, damit die Selbst-Prozessierung noch gezielter gestartet werden kann. Zudem würde der Zusatz eines Inhibitors den Ausbeuteverlust an unprozessierter RNA bei der *in vitro* Transkription beschränken.

4. Literaturverzeichnis

1. Little, J.W. (1993) LexA cleavage and other self-processing reactions. *J Bacteriol*, **175**, 4943-4950.
2. Kuroyanagi, M., Nishimura, M. and Hara-Nishimura, I. (2002) Activation of Arabidopsis vacuolar processing enzyme by self-catalytic removal of an auto-inhibitory domain of the C-terminal propeptide. *Plant Cell Physiol*, **43**, 143-151.
3. Lee, J., O'Hare, T., Pilch, P.F. and Shoelson, S.E. (1993) Insulin receptor autophosphorylation occurs asymmetrically. *J Biol Chem*, **268**, 4092-4098.
4. Shoelson, S.E., Boni-Schnetzler, M., Pilch, P.F. and Kahn, C.R. (1991) Autophosphorylation within insulin receptor beta-subunits can occur as an intramolecular process. *Biochemistry*, **30**, 7740-7746.
5. Kruger, K., Grabowski, P.J., Zaugg, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, **31**, 147-157.
6. Dallas, A., Balatskaya, S.V., Kuo, T.-C., Ilves, H., Vlassov, A.V., Kaspar, R.L., Kisich, K.O., Kazakov, S.A. and Johnston, B.H. (2008) Hairpin ribozyme-antisense RNA constructs can act as molecular lassos. *Nucleic Acids Research*, **36**, 6752-6766.
7. Gilbert, W. (1986) Origin of life: The RNA world. *Nature*, **319**, 618-618.
8. Diener, T.O. (1989) Circular RNAs: relics of precellular evolution? *Proc Natl Acad Sci U S A*, **86**, 9370-9374.
9. Ma, W., Yu, C. and Zhang, W. (2013) Circularity and self-cleavage as a strategy for the emergence of a chromosome in the RNA-based protocell. *Biology Direct*.
10. van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F. and Grivell, L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, **44**, 225-234.
11. Cocquerelle, C., Mascrez, B., Hetuin, D. and Bailleul, B. (1993) Mis-splicing yields circular RNA molecules. *FASEB J*, **7**, 155-160.
12. Zaphiropoulos, P.G. (1996) Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc Natl Acad Sci U S A*, **93**, 6536-6541.
13. Zhang, Y., Zhang, X.O., Chen, T., Xiang, J.F., Yin, Q.F., Xing, Y.H., Zhu, S., Yang, L. and Chen, L.L. (2013) Circular intronic long noncoding RNAs. *Mol Cell*, **51**, 792-806.
14. Murray, H.L., Mikheeva, S., Coljee, V.W., Turczyk, B.M., Donahue, W.F., Bar-Shalom, A. and Jarrell, K.A. (2001) Excision of group II introns as circles. *Mol Cell*, **8**, 201-211.
15. Bak, R.O. and Mikkelsen, J.G. (2013) miRNA sponges: soaking up miRNAs for regulation of gene expression. *Wiley Interdiscip Rev RNA*.
16. Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384-388.
17. Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. and Sharpless, N.E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, **19**, 141-157.
18. Liu, Y., Cui, H., Wang, W., Li, L., Wang, Z., Yang, S. and Zhang, X. (2013) Construction of circular miRNA sponges targeting miR-21 or miR-221 and demonstration of their excellent anticancer effects on malignant melanoma cells. *Int J Biochem Cell Biol*, **45**, 2643-2650.
19. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M. et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333-338.
20. Wang, P.L., Bao, Y., Yee, M.C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinneny, J.R., Brown, P.O. and Salzman, J. (2014) Circular RNA Is Expressed across the Eukaryotic Tree of Life. *PLoS One*, **9**, e90859.
21. Lukiw, W.J. (2013) Circular RNA (circRNA) in Alzheimer's disease (AD). *Front Genet*, **4**, 307.

22. Ghosal, S., Das, S., Sen, R., Basak, P. and Chakrabarti, J. (2013) Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front Genet*, **4**, 283.
23. Kos, A., Dijkema, R., Arnberg, A.C., van der Meide, P.H. and Schellekens, H. (1986) The hepatitis delta (delta) virus possesses a circular RNA. *Nature*, **323**, 558-560.
24. Flores, R., Grubb, D., Elleuch, A., Nohales, M.A., Delgado, S. and Gago, S. (2011) Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. *RNA Biol*, **8**, 200-206.
25. Sogo, J.M. and Schneider, I.R. (1982) Electron microscopy of double-stranded nucleic acids found in tissue infected with the satellite of tobacco ringspot virus. *Virology*, **117**, 401-415.
26. Hampel, A. and Tritz, R. (1989) RNA catalytic properties of the minimum (-)sTRSV sequence. *Biochemistry*, **28**, 4929-4933.
27. Esteban, J.A., Walter, N.G., Kotzorek, G., Heckman, J.E. and Burke, J.M. (1998) Structural basis for heterogeneous kinetics: reengineering the hairpin ribozyme. *Proc Natl Acad Sci U S A*, **95**, 6091-6096.
28. Shepotinovskaya, I.V. and Uhlenbeck, O.C. (2008) Catalytic diversity of extended hammerhead ribozymes. *Biochemistry*, **47**, 7034-7042.
29. Tuerk, C., Gauss, P., Thermes, C., Groebe, D.R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O.C., Tinoco, I., Jr. et al. (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci U S A*, **85**, 1364-1368.
30. Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M. and Uesugi, S. (1990) Studies on the structure and stabilizing factor of the CUUCGG hairpin RNA using chemically synthesized oligonucleotides. *Nucleic Acids Res*, **18**, 3831-3839.
31. Berzal-Herranz, A., Joseph, S. and Burke, J.M. (1992) In vitro selection of active hairpin ribozymes by sequential RNA-catalyzed cleavage and ligation reactions. *Genes Dev*, **6**, 129-134.
32. Walter, N.G., Hampel, K.J., Brown, K.M. and Burke, J.M. (1998) Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *EMBO J*, **17**, 2378-2391.
33. Petkovic, S. and Muller, S. (2013) RNA self-processing: Formation of cyclic species and concatemers from a small engineered RNA. *FEBS Lett*, **587**, 2435-2440.
34. Ryder, S.P. and Strobel, S.A. (1999) Nucleotide analog interference mapping of the hairpin ribozyme: implications for secondary and tertiary structure formation. *J Mol Biol*, **291**, 295-311.
35. Chowrira, B.M., Berzal-Herranz, A., Keller, C.F. and Burke, J.M. (1993) Four ribose 2'-hydroxyl groups essential for catalytic function of the hairpin ribozyme. *J Biol Chem*, **268**, 19458-19462.
36. Rupert, P.B. and Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780-786.
37. Pinard, R., Lambert, D., Walter, N.G., Heckman, J.E., Major, F. and Burke, J.M. (1999) Structural basis for the guanosine requirement of the hairpin ribozyme. *Biochemistry*, **38**, 16035-16039.
38. Mlynksy, V., Banas, P., Walter, N.G., Sponer, J. and Otyepka, M. (2011) QM/MM studies of hairpin ribozyme self-cleavage suggest the feasibility of multiple competing reaction mechanisms. *J Phys Chem B*, **115**, 13911-13924.
39. Kath-Schorr, S., Wilson, T.J., Li, N.S., Lu, J., Piccirilli, J.A. and Lilley, D.M. (2012) General acid-base catalysis mediated by nucleobases in the hairpin ribozyme. *J Am Chem Soc*, **134**, 16717-16724.
40. Ferre-D'Amare, A.R. (2004) The hairpin ribozyme. *Biopolymers*, **73**, 71-78.
41. Sumita, M., White, N.A., Julien, K.R. and Hoogstraten, C.G. (2013) Intermolecular domain docking in the hairpin ribozyme: metal dependence, binding kinetics and catalysis. *RNA Biol*, **10**, 425-435.

42. Klostermeier, D. and Millar, D.P. (2002) Energetics of hydrogen bond networks in RNA: hydrogen bonds surrounding G+1 and U42 are the major determinants for the tertiary structure stability of the hairpin ribozyme. *Biochemistry*, **41**, 14095-14102.
43. Wilson, T.J., Nahas, M., Araki, L., Harusawa, S., Ha, T. and Lilley, D.M. (2007) RNA folding and the origins of catalytic activity in the hairpin ribozyme. *Blood Cells Mol Dis*, **38**, 8-14.
44. Wilson, T.J. and Lilley, D.M. (2011) Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data. *RNA*, **17**, 213-221.
45. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
46. Flamm, C., Hofacker, I.L., Maurer-Stroh, S., Stadler, P.F. and Zehl, M. (2001) Design of multistable RNA molecules. *RNA*, **7**, 254-265.
47. Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**, 26.
48. Haseloff, J. and Gerlach, W.L. (1989) Sequences required for self-catalysed cleavage of the satellite RNA of tobacco ringspot virus. *Gene*, **82**, 43-52.
49. Feldstein, P.A., Buzayan, J.M. and Bruening, G. (1989) Two sequences participating in the autolytic processing of satellite tobacco ringspot virus complementary RNA. *Gene*, **82**, 53-61.
50. Buzayan, J.M., Gerlach, W.L. and Bruening, G. (1986) Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *Proc Natl Acad Sci U S A*, **83**, 8859-8862.
51. Rodbard, D. and Chrambach, A. (1971) Estimation of Molecular Radius, Free Mobility, and Valence Using Polyacrylamide Gel Electrophoresis. *Analytical Biochemistry*, **40**, 95-134.
52. Gunnarsson, G.H., Guðmundsson, B., Thormar, H.G., Alfredsson, A. and Jonsson, J.J. (2006) Two-dimensional strandness-dependent electrophoresis. *Nat Protoc*, **1**, 3011-3018.
53. Umekage, S. and Kikuchi, Y. (2009) In vitro and in vivo production and purification of circular RNA aptamer. *J Biotechnol*, **139**, 265-272.
54. Mattila, P., Korpela, J., Tenkanen, T. and Pitkanen, K. (1991) Fidelity of DNA synthesis by the Thermococcus litoralis DNA polymerase--an extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res*, **19**, 4967-4973.
55. Fay, M.J., Walter, N.G. and Burke, J.M. (2001) Imaging of single hairpin ribozymes in solution by atomic force microscopy. *RNA*, **7**, 887-895.
56. Bernhart, S.H., Tafer, H., Muckstein, U., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol*, **1**, 3.
57. Gan, H.H., Fera, D., Zorn, J., Shiffeldrim, N., Tang, M., Laserson, U., Kim, N. and Schlick, T. (2004) RAG: RNA-As-Graphs database--concepts, analysis, and features. *Bioinformatics*, **20**, 1285-1291.
58. Cheong, C., Varani, G. and Tinoco, I., Jr. (1990) Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature*, **346**, 680-682.

Critical Review

The Many Faces of the Hairpin Ribozyme: Structural and Functional Variants of a Small Catalytic RNA

Sabine Müller, Bettina Appel, Tobias Krellenberg, and Sonja Petkovic

Ernst-Moritz-Arndt Universität Greifswald, Institut für Biochemie, Felix-Hausdorff Str. 4, 17487 Greifswald, Germany

Summary

The hairpin ribozyme is a small catalytic RNA that has been reengineered resulting in a number of variants with extended or even new functions. Thus, manipulation of the hairpin ribozyme structure has allowed for activity control by external effectors, namely oligonucleotides, flavine mononucleotide, and adenine. Hairpin ribozyme-derived twin ribozymes that mediate RNA fragment exchange reactions as well as self-processing hairpin ribozymes were designed. Furthermore, several hairpin ribozyme variants have been engineered for knock down of specific RNA substrates by adapting the substrate-binding domain to the specific target sequence. This review will focus on hairpin ribozymes possessing structural extensions/variations and thus functionally differing from the parent hairpin ribozyme. © 2011

IUBMB

IUBMB Life, 64(1): 36–47, 2012

Keywords ribozyme; oligonucleotides; extended linker; Watson–Crick base pair; phosphodiester linkage.

Abbreviations TRAP, trp-RNA-binding attenuation protein; FMN, flavine mononucleotide; ADHR, adenine-dependent hairpin ribozymes; MAPK, mitogen-activated protein kinase.

INTRODUCTION

Over the past 20 years, a large number of ribozymes, riboswitches, and other RNAs with specific functions have been discovered in nature. The structure and the mechanism of those RNAs have been extensively studied and are well understood in many cases, nowadays. The knowledge of the working mecha-

nism of ribozymes and riboswitches has allowed designing RNAs as molecular tools with potential applications, for example, in molecular biology, medicine, and diagnostics (1–4). Moreover, the powerful approach of in vitro selection has created a large variety of novel RNA catalysts with activities that so far have not been identified in nature. In addition to in vitro selection, also the construction of catalytic RNAs with new or at least extended functions by rational design has become a major topic.

Among the small RNAs occurring in nature, the hammerhead and the hairpin ribozymes are the most studied and varied structures. Both RNA motifs are sufficiently small to be efficiently prepared by chemical synthesis, facilitating their specific modifications for structure and functional studies and for RNA engineering. The hairpin ribozyme (Fig. 1) catalyzes the reversible cleavage of a specific phosphodiester bond of an appropriate RNA substrate, generating characteristic fragments, one with 2',3'-cyclic phosphate and the other with a 5'-OH terminus (5, 6).

It is derived from the minus strand of the satellite RNA from tobacco ringspot virus, where it is responsible for cleavage and likewise ligation of intermediates occurring during rolling circle replication of the viral satellite RNA (7). The cleavage reaction proceeds via in-line attack of the 2'-hydroxy group on the adjacent phosphorous atom, leading to the departure of the 5'-oxygen atom on the adjacent ribose, and to generation of the 2',3'-cyclic phosphate (8). Typical rates of multiple turnover reactions are between 0.01 and 1 molecule per minute at physiological salt concentrations, consistent with a substantial 10⁷–10⁹-fold rate enhancement over uncatalyzed non-specific RNA cleavage. Also the reverse reaction is catalyzed. In this case, the 5'-oxygen atom attacks the phosphorus of the 2',3'-cyclic phosphate group leading to ring opening and consequently to ligation. The hairpin ribozyme is particularly efficient in catalyzing the ligation reaction. Strikingly, ligation is favored if the hairpin ribozyme is folded into a stable three-dimensional structure, while cleavage occurs from ribozyme–substrate complexes that are less stable, yet stable enough to fold into a catalytically competent structure (9). The secondary structure of the minimal catalytic motif consists of four Watson–Crick base paired

Received 1 July 2011; accepted 18 August 2011

Address correspondence to: Sabine Müller; Ernst-Moritz-Arndt Universität Greifswald, Institut für Biochemie, Felix-Hausdorff Str. 4, 17487 Greifswald, Germany. Tel: +49 3834 8622843. Fax: +49 3834 864471. E-mail: smueller@uni-greifswald.de

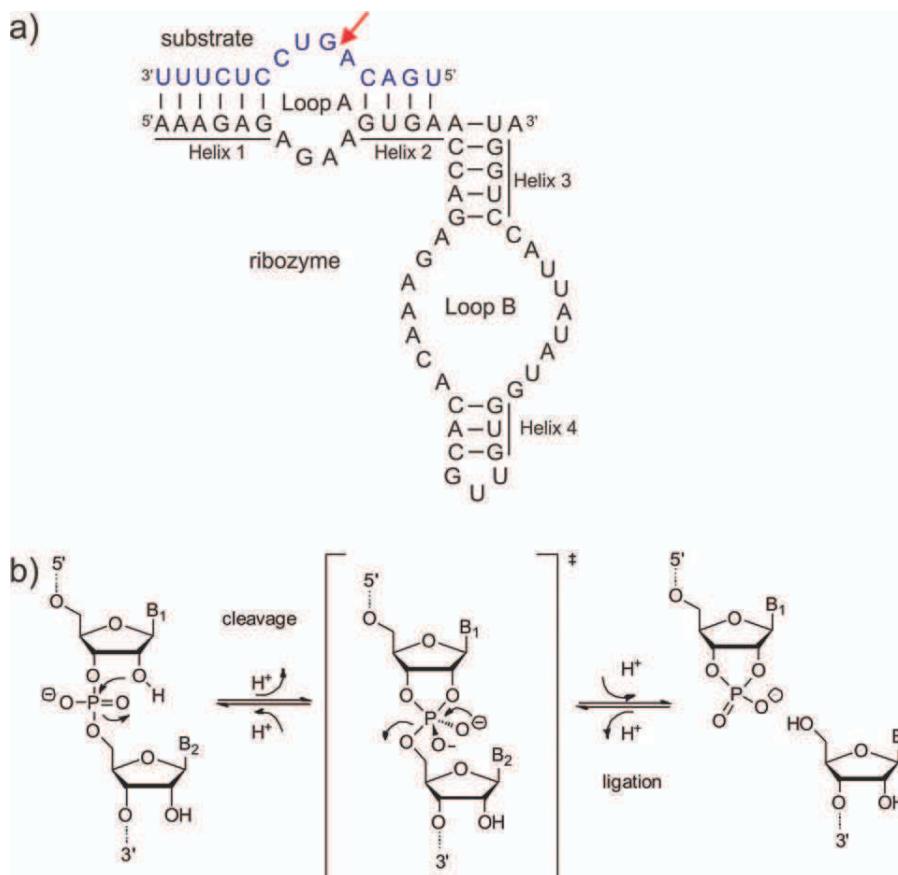


Figure 1. a) Secondary structure of the hairpin ribozyme. The arrow denotes the site of reversible cleavage. b) Mechanism of reversible phosphodiester cleavage. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

helices separated by two internal loops A and B (Fig. 1). In the viral satellite RNA, the hairpin ribozyme is embedded in a four way junction, which assists tertiary folding. Intimate association of loops A and B, called docking, is stabilized by divalent metal ions and generates the local environment in which catalysis occurs. The docked ribozyme is further stabilized by a Watson-Crick base pair between a loop A guanosine and a loop B cytidine as well as a ribose zipper involving nucleotides of both loops (10). A minimal version of the ribozyme in which the four-way junction is replaced by a simple phosphodiester linkage is active, but requires a one thousand fold higher Mg^{2+} concentration to induce folding (11). Upon binding to its substrate, the hairpin ribozyme undergoes extensive structural rearrangement, which results in distortion of the substrate RNA that primes it for cleavage. To achieve catalysis, the ribozyme is likely employing exclusively RNA functional groups. The crystal structure of a precursor form of the RNA as well as biochemical experiments have suggested that an active-site adenine and an active site guanosine might adopt the role of a general acid or base (12, 13). Metal ions or other catalytic cofactors are not used for active site chemistry.

Among the many hairpin ribozyme variants, a number of species were designed for knock down of specific RNA substrates by adapting the substrate-binding domain to the specific target sequence (14). According to this purpose, there is also some work concerning the development of hairpin ribozymes for poorly processed substrates (15, 16). These sequence variants of the hairpin ribozyme will not be considered in this review. Herein we focus on hairpin ribozymes possessing structural extensions/variations and thus being functionally different compared with the parent hairpin ribozyme.

STRUCTURAL VARIANTS OF THE HAIRPIN RIBOZYME

A strong motivation for the design of variants of a certain biomolecule is the aim to study its structure and function. This often requires making the molecule prone for a certain analytical method. The first 2.4 Å resolution structure of a hairpin ribozyme was successfully solved by crystallization of a four-helix junction form of the hairpin ribozyme in complex with a substrate strand in which the nucleophilic 2'-OH group was replaced with a methoxy function, and most importantly, with

the basic RNA-binding protein U1A (10). Thereby, the cognate site for U1A was grafted onto the structurally and functionally dispensable distal end of stem B. Crystallization of the RNA together with the basic protein was facilitated by structural stabilization of the RNA (*i*) by the four-way junction structure and (*ii*) by the protein. A ribonucleoprotein variant of the hairpin ribozyme was also used as a model system to study biological catalysis by ribonucleoproteins. Here, the hairpin ribozyme structure was modified by inserting a binding site for bacteriophage R17 coat protein in the form of an extension to helix 4, located sufficiently apart from the catalytic domain (17). The extended construct was found being more efficient in catalysis of the cleavage reaction (2-fold) and even more in catalysis of the ligation reaction (16-fold) compared with the parent wild-type hairpin ribozyme. However, both reactions were found being insensitive to binding of the protein.

Most biochemical experiments have been carried out using constructs where the two stem-loop domains are connected by a single stranded linker. However, in the viral satellite RNA, stem-loop domains A and B are part of a four-helix junction, and FRET measurements demonstrated that docking of loops A and B, required for catalysis, is greatly favored in constructs that contain the four-helix junction compared with minimal hairpin ribozymes, where the two stems are connected by an extended linker (18–20). In addition to four-helix junction structures, also hairpin ribozymes organized in three-way junctions were used. These constructs have a slightly stabilized structure compared with the two arm hairpin ribozymes, and thus are more efficient in catalysis of both cleavage and ligation (15, 21–23). As shown by Burke and co-workers, three-way junction hairpin ribozymes support folding in the active structure with the two loops A and B in close proximity (23). This, in turn, disfavors coaxial stacking of the two ribozyme arms containing loop A and B, as observed for the minimal two-helix form of the hairpin ribozyme. A hairpin ribozyme variant having a non-nucleotidic orthobenzene interface that connects the two domains was shown having similar supportive effects on folding as the three-way junction hairpin ribozyme structures (23).

Another type of three-way junction hairpin ribozymes was engineered by Ohtsuka and co-workers (24). With the aim of trimming the hairpin ribozyme construct after T7 transcription from a DNA template, another domain A' was connected to the 3'-end of domain B of the parent hairpin ribozyme, thus generating a ribozyme variant with three-stem-loop domains (Fig. 2). While domain A binds the appropriate RNA substrate for *trans*-cleavage, domain A' possesses *cis*-cleavage activity leading to self-trimming. The sequence of domain A' was derived from domain A, such that domain B acts as catalytic center for both *trans*-cleavage and *cis*-cleavage.

As shown by Gait and co-workers (25, 26) and by our laboratory (27), efficient hairpin ribozyme cleavage can be affected by millimolar concentrations of the tetraamine spermine, even in the absence of metal ions. This observation later led to the synthesis of hairpin ribozyme–spermine conjugates with the

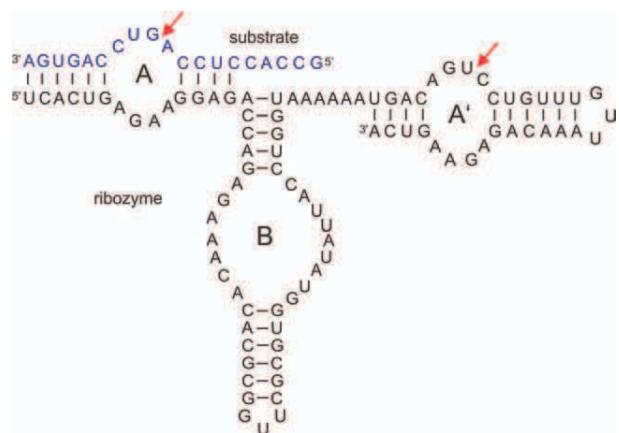


Figure 2. Hairpin ribozyme construct with three stem-loop domains (24). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

polyamine covalently attached to the C5 position of specific uridines in loop B and helix 4 (27). However, while the modification in most cases was well tolerated by the ribozyme, the requirement for magnesium ion cofactors for all tested spermine-modified ribozymes was not decreased compared to the wild-type ribozyme.

In the earlier days of hairpin ribozyme research, where no crystal structure was yet available, a number of biochemical studies were carried out to find evidence for the interaction of loop A with loop B. As was already known, essential nucleotides of the ribozyme reside within the two loops, and linker insertion (28, 29) as well as chemical modification studies (30) of the stem 2–3 junction indicated that stem 2 is not coaxially stacked upon stem 3 in the catalytically competent structure. This allowed for the conclusion that a bend between stems 2 and 3 is required to bring essential elements in loop B proximal to the cleavage site located in loop A. Further support came from the study of hairpin ribozyme variants, where the two domains were physically separated (31, 32). The resulting constructs rely on long-range tertiary contacts for catalysis. Burke and co-workers used an optimized ribozyme and substrate, which included sequence changes to minimize the formation of nonproductive conformational isomers (31). The overall catalytic efficiency of the reconstituted ribozyme was found being about 100-fold lower compared with the intact ribozyme. However, the decrease resulted entirely from a 100-fold increase in the apparent K_M , whereas k_{cat} remained virtually unchanged, demonstrating that cleavage chemistry is not impaired by the topology of the ribozyme, but that reaction is limited by productive assembly of the two domains. Similar results were obtained by Shin et al. (32).

Non-hinged hairpin ribozymes were also used in different context by Vlassov et al. (33). In a study focussing on evolution of complex RNAs in the RNA world, the authors investigated the ligation of RNA substrates by truncated and fragmented

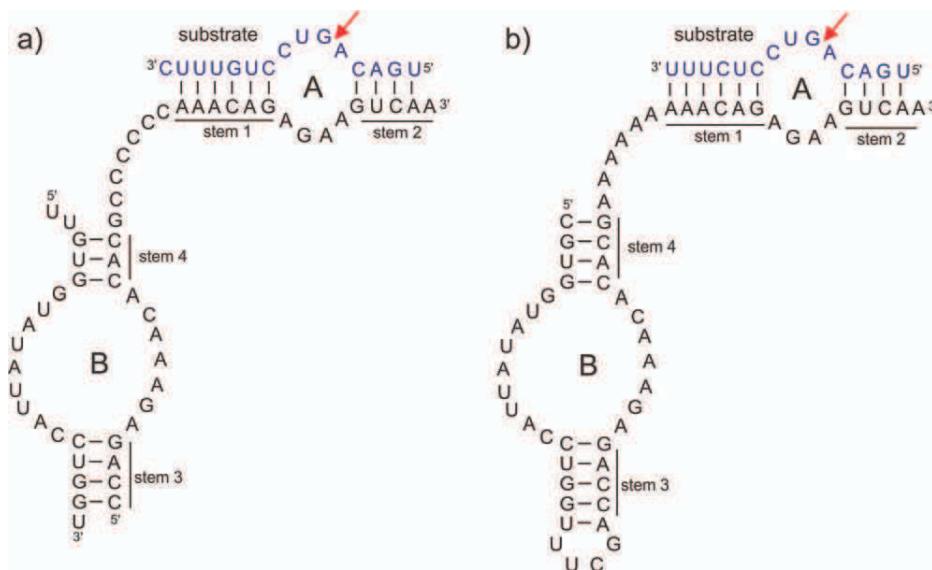


Figure 3. Hairpin ribozymes with reverse-joined domains. Stem 4 is connected in reverse order to stem 1 by an oligocytidine linker (34) (a) or an oligoadenosine linker (35, 36) (b). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

derivatives of the hairpin ribozyme at low temperatures. Several of the truncated and fragmented constructs indeed were found being capable of ligation of a wide variety of RNA molecules to a given sequence in frozen solution, whereas under standard conditions little or no activity was observed.

Another line of experiments looking at hairpin ribozyme variants with reverse-joined domains gave further evidence for the importance and strength of the tertiary interdomain interaction. In these constructs, domains were also separated, but then rejoined in reverse order (Fig. 3). A hairpin ribozyme was divided at the hinge region between stem 2 and 3, and the loop B domain was rejoined via stem 4 to stem 1 of the loop A domain by oligocytidine (34) or oligoadenosine (35, 36) linkers of different length to maintain the interaction of the domains required for catalysis.

Dependent on the linker length, the reverse-joined constructs retained cleavage (34, 35) and even ligation (36) activity. A special kind of reverse-joined hairpin ribozymes was engineered in our lab (37) (Fig. 4). In this construct, 2'-deoxy-N4-(6-hydroxyhexyl)-5-methylcytidine was introduced to connect the loop B domain with the loop A domain via an artificial branch. The synthesized branched RNA was found to cleave its substrate only 5-fold lower compared with the corresponding nonbranched reverse-joined ribozyme. However, no ligation activity could be detected, which presumably is the result of insufficient stability of the catalytic competent conformation. We have suggested that cleavage might occur only from temporarily folded species that are insufficiently stable to allow for ligation (37).

Apart from their meaning for the characterization of the catalytic topology of the hairpin ribozyme, reverse-joined hairpin ribozyme variants have been used as functional modules in twin

ribozymes for cleavage (35) and repair (36) of small deletions in RNA substrates (see below).

TWIN RIBOZYMES

Twin ribozymes are engineered RNA structures that mediate two chain cleavages and two ligation events in a strictly controlled fashion (36, 38–40) (Fig. 5). They are derived from the hairpin ribozyme by tandem duplication (38–40) or by combination of a conventional hairpin ribozyme unit with a reverse-joined hairpin ribozyme (35, 36) (see above). Initially, twin ribozymes were engineered for RNA double cleavage (35). However, the specific cleavage/ligation properties of hairpin ribozyme derived twin ribozymes can be used to cleave a pre-defined sequence patch out of a suitable RNA substrate followed by ligating a separate synthetic “repair” fragment into the gap left behind. The key element in this process is to engineer the ribozyme in a manner that allows easy and fast dissociation of the cleaved out fragment versus tight binding of the repair oligonucleotide. Ligation is enthalpically favored, and the entropic cost for hairpin ribozyme ligation is rather low (41). Thus, fragments that are tightly bound to the twin ribozyme become preferentially ligated, while fragments that are less stable can easily dissociate from the ribozyme and thus are preferentially cleaved (38). Considering these principles, dependent on the specific design of the twin ribozyme–substrate complex, fragments of equal lengths are exchanged (39), or a shorter patch is replaced with a longer one (36, 38, 39) and vice versa (40) (Fig. 5). Thus, the system mimics the repair of short deletions, insertions, and base replacement mutations with up to 53% yield (39).

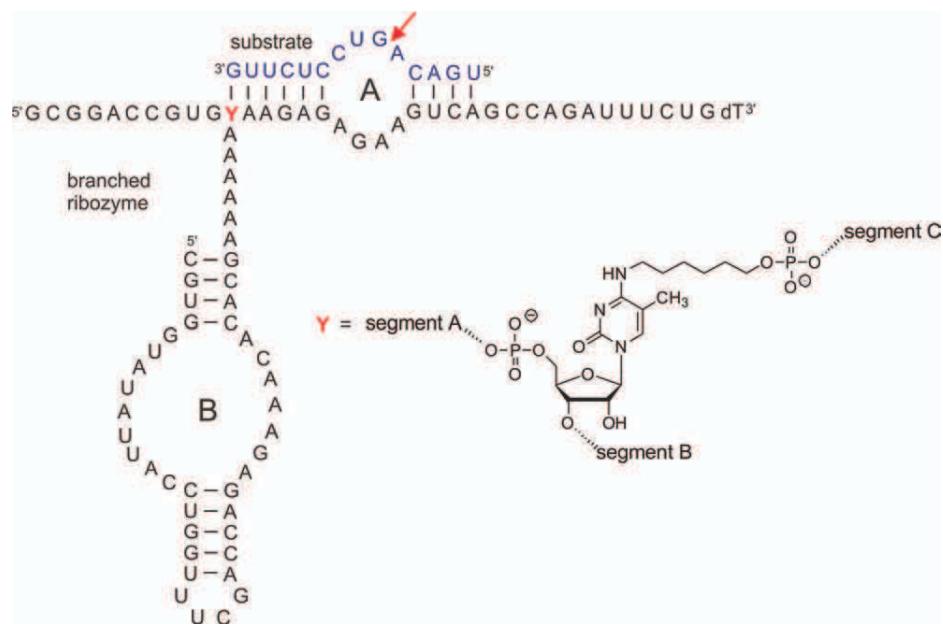


Figure 4. Branched hairpin ribozyme. The structure of the branch point nucleoside is shown separately (37). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

To demonstrate the potential of twin ribozymes for functional repair of genetically relevant substrates, we have developed a twin ribozyme targeting the mutated transcript of the *CTNNB1* gene, encoding β -catenin, a major player of the wnt-signaling pathway (42, 43). Recently, *in vitro* repair could be successfully demonstrated (44). Apart from the therapeutic potential, twin ribozymes may have played a role as recombinases in the RNA world. Ribozyme-mediated catalysis of RNA recombination may have contributed to extending the sequence space and the function of RNA molecules in early life forms.

HAIRPIN RIBOZYMES THAT ARE REGULATED BY EXTERNAL EFFECTORS

Over the past decade, a large number of RNA structures have been engineered being composed of a catalytic domain joined to a specific aptamer via a communication module (45). This allows for allosteric control of ribozyme activity: dependent on the presence of the specific ligand, activity is switched ON or OFF, or is at least up or down regulated. In this way, the ligand acts as external effector or inhibitor. However, not all systems correspond to the principle of allosteric regulation in the classical way, which would require the allosteric cofactor binding to a region apart from the catalytic domain, and inducing a conformational change in the binding region that is transduced to the catalytic domain. In particular, when oligonucleotides are used for activity regulation of ribozymes, other strategies like inhibitor control, complementation, trap, or expansive regulation strategies have been used (46).

There are several hairpin ribozyme constructs that are regulated by external oligonucleotides (47–51). One of the first examples is a hairpin variant obtained by *in vitro* selection that

is cleavage active in the presence of an exogenous short oligonucleotide (47). Thereby, the short oligonucleotide effector presumably binds to a complementary sequence located in a hairpin loop closing helix 4. It was suggested that in the absence of the oligonucleotide, the loop domain forms a slipped hairpin loop, putting the binding sequence for the regulatory oligonucleotide in a single stranded region. Upon binding of the oligonucleotide, a structural change of the ribozyme is triggered leading to a catalytically competent fold.

A different scenario was used in our laboratory (48). Instead of introducing an additional allosteric binding site, we designed a hairpin ribozyme derivative that requires an oligonucleotide effector as a structural element for the formation of the catalytic center. This system is not truly allosteric, it rather represents an example for ribozyme activation by complementation. Addition of an oligonucleotide restores the catalytic center, and thus activity of a formerly inactive ribozyme. Inactivation was achieved by mutating C25 to G, as well as by introduction of G–U wobble base pairs and mismatches in helix 4 for structural destabilization (Fig. 6).

The oligonucleotide effector contains the essential G25, and moreover, is complementary to the sequences in stem 3 and 4 of the inactive hairpin ribozyme variants. The effector was shown being capable of invading minimal hinged and three-way junction hairpin ribozyme structures (Fig. 6), thereby restoring stem 3, loop B, and stem 4 as required for activity. Compared to the corresponding wild-type structures, both the minimal hinged and the three-way junction hairpin ribozyme variant, in the presence of an 5- to 8-fold excess of the effector oligonucleotide, cleaved their substrate nearly to completion (96%), although with about 5-fold lower rates (48).

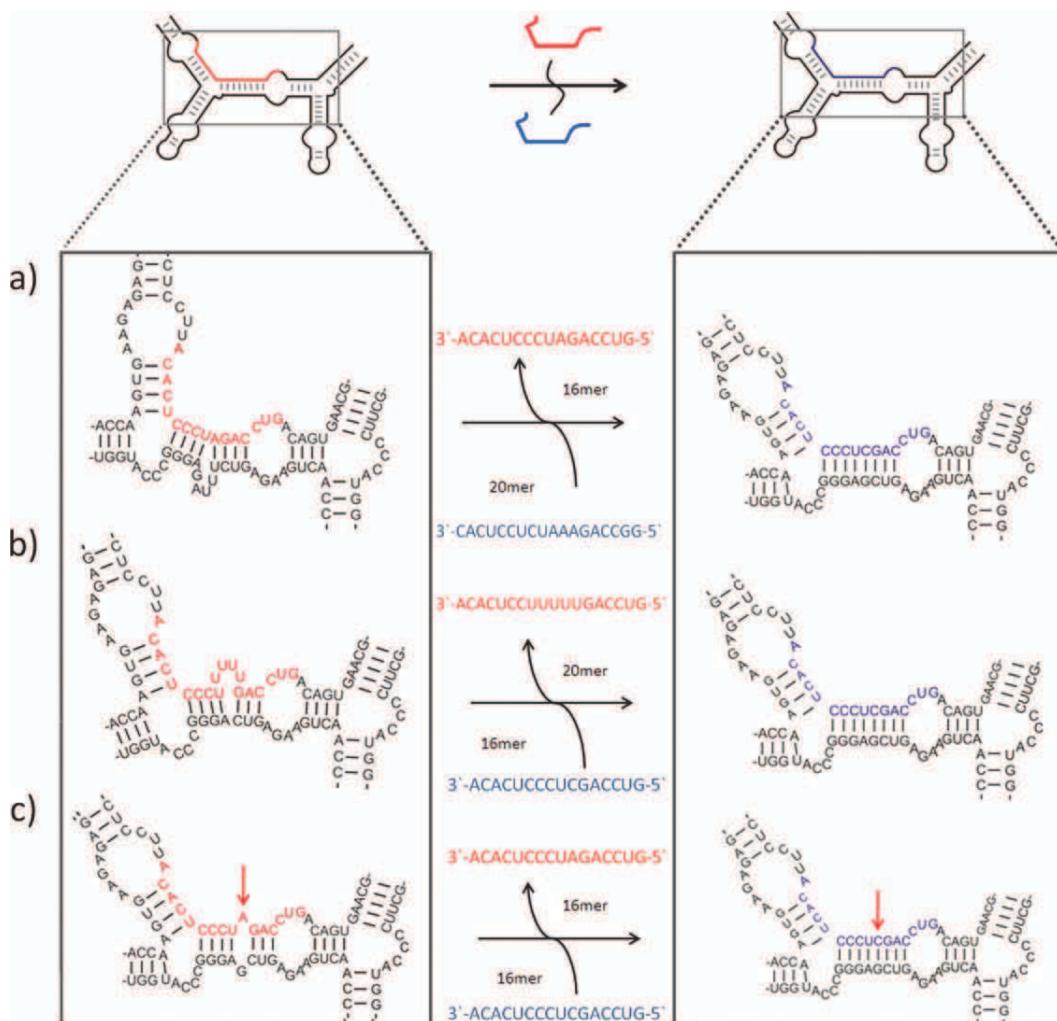


Figure 5. Twin ribozyme mediated fragment exchange. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Famulok and co-workers designed hairpin ribozyme variants that can be induced or repressed by external effector oligonucleotides by interfering with the docking process that brings the two domains A and B in close proximity as required for activity (49). Najafi-Shoushtari et al. incorporated a third domain C via a pseudo-three way junction to assimilate alternate stable RNA motifs such as an internal stem-loop structure in rHP-TRAP (Fig. 7a) or a pseudo-half-knot in iHP-TRAP (Fig. 7b).

Small sequence variations in domain C allowed targeted switching of ribozyme activity, and it was demonstrated that the same effector oligonucleotide can either serve as an inducer or repressor. In this assay, instead of a short oligonucleotide a specific mRNA was used as effector. The hairpin ribozyme variants were applied to *trp* leader mRNA, which is the RNA sequence that is bound by L-tryptophan-activated *trp*-RNA-binding attenuation protein (TRAP). When domain C contained a sequence patch complementary to *trp* leader mRNA, activity of

the ribozyme variants was shown to be altered by annealing of the *trp*-mRNA. The effect could be reverted by TRAP/tryptophan-mediated sequestration of the ribozyme-binding site in the *trp*-mRNA. In particular, hybridization of *trp*-mRNA to the repressible hairpin ribozyme variant rHP-TRAP inhibits catalysis by forcing domains A and B into an extended conformation (Fig. 7a). Used with the inducible hairpin ribozyme variant, the same *trp*-mRNA triggers a conformational change in iHP-TRAP generating a pseudo-half knot structure that is stabilized by the terminal helical sequence at the pseudo-three-way junction, thereby switching activity ON. Maximal rate enhancement was found at a ratio of 1:1 of *trp*-mRNA and rHP-TRAP, or iHP-TRAP (49). Similarly, a hairpin ribozyme variant was constructed that can be regulated by a short DNA aptamer specific for human α -thrombin (51). Najafi-Shoushtari und Famulok have made use of the principle of applying two regulatory factors that switch ribozyme activity in opposite directions in

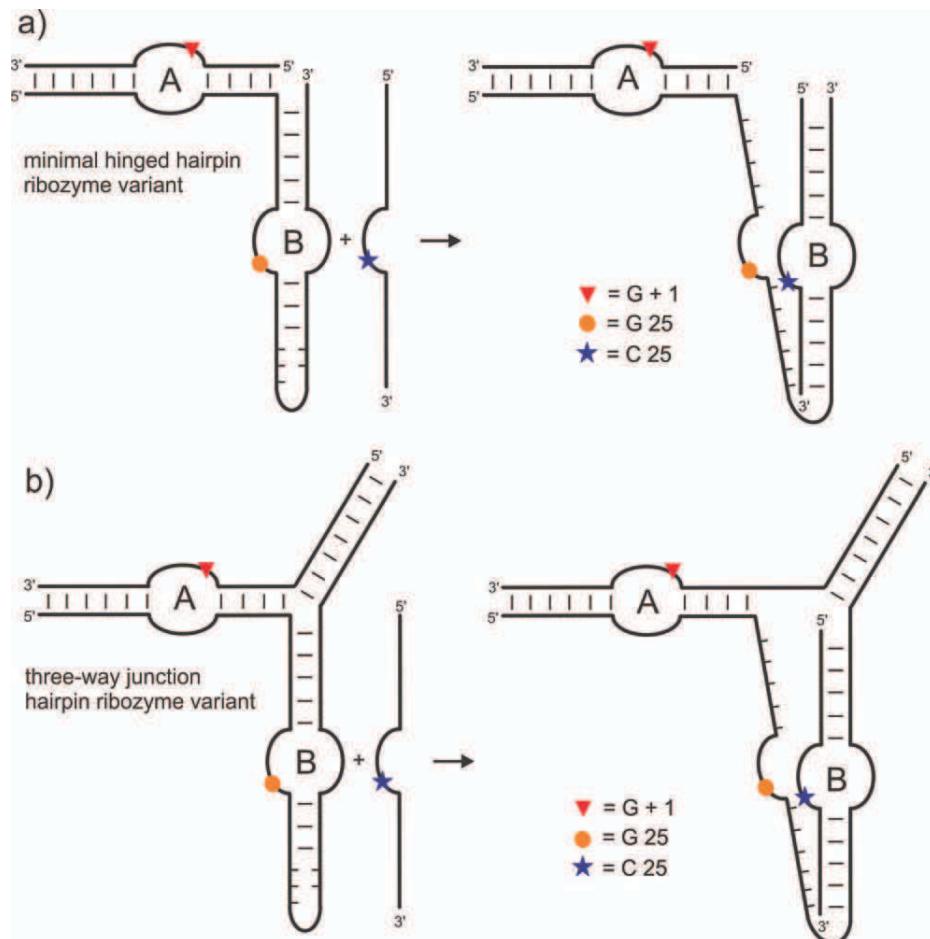


Figure 6. Complementation of inactive hairpin ribozyme structures by an oligonucleotide effector (48). The minimal hinged form (a) and a three-way junction hairpin ribozyme (b) were used. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

further hairpin ribozyme engineering. They constructed a variant that in domain C harbors an aptamer sequence that binds flavine mononucleotide (FMN) (50). Formation of the FMN–aptamer complex facilitates docking of domains A and B, thus inducing activity. A short oligonucleotide, that is complementary to the aptamer, abolishes ribozyme activity by formation of an extended conformer unable to perform catalysis. Addition of FMN to the inactive oligonucleotide-ribozyme complex was found to competitively neutralize the inhibitory effect of the oligonucleotide and to activate the ribozyme.

FMN responsive hairpin ribozyme variants have been engineered also in our laboratory (52, 53). In our constructs, stem 4 was replaced with two different sequences acting as communication modules and connecting the ribozyme moiety with the FMN-specific aptamer (HPAR2 and HPAR5 in Fig. 8a). Both variants were found being regulated by FMN, although the increase of activity brought about by binding of FMN was only 24- and 13-fold. However, this activity increase was sufficient to observe regulation in dependence of the oxidation state of

FMN. Binding of FMN to its aptamer to a large extent is stabilized by hydrophobic stacking of the aromatic isoalloxazin ring with the bases of the aptamer (54). Thus, FMN binding could be reverted by addition of the reducing agent $\text{Na}_2\text{S}_2\text{O}_4$. Upon reduction, the molecular shape of FMN changes from planar to roof-like, leading to loss of its binding capacity, and subsequently to down regulation of activity (52) (Fig. 8b).

Iterative cycles of oxidation/reduction allow for reversible switching of activity. Instead of the chemical $\text{Na}_2\text{S}_2\text{O}_4$, also electrochemical reduction/oxidation was used for control of FMN binding (53). Currently we work on the development of hairpin ribozyme variants that are repressed in the presence of FMN (55). Based on computational selection studies, another FMN-dependent hairpin ribozyme variant was developed, wherein the FMN aptamer was appended to the end of loop B via a symmetrical region of three base pairs (56). This variant was found to be activated 10-fold in the presence of FMN.

In contrast to the FMN-dependent hairpin ribozyme variants described above, where the specific FMN aptamer was conjugated

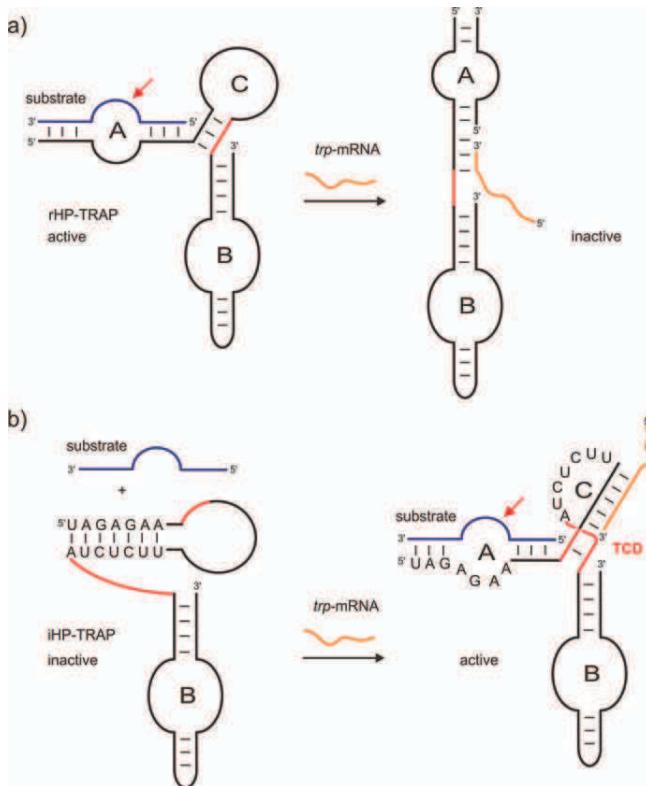


Figure 7. Repressible (a) and inducible (b) hairpin ribozyme variants. Activity is dependent on a specific RNA sequence (*trp*-mRNA) that interrupts (a) or stabilizes (b) the active conformation (49). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

to the hairpin ribozyme helix 4, Maurel and co-workers have developed adenine-dependent hairpin ribozymes (ADHR) with adenine presumably being bound in the ribozyme structure itself (57–61). The exact binding mode and binding site were analyzed by NMR spectroscopy (62). However, no specific ligand binding or low-affinity RNA–ligand complexes could be detected, suggesting that the interaction of the ribozyme with the adenine cofactor is only transient or at least too weak to significantly change the structure of the ribozyme. The adenine-dependent hairpin ribozyme variants were selected *in vitro*. The secondary structure consists of the minimal wild-type hairpin ribozyme elements with four helical regions that are separated by two loop regions. The two domains are connected by an oligocytidine linker. Both adenine-dependent hairpin ribozymes ADHR1 and ADHR2 found by *in vitro* selection have Mg²⁺ requirements for catalysis similar to the wild-type hairpin ribozyme. Based on the investigation of pH effects on catalytic rates, it was suggested that ADHRs use new acid/base catalytic strategies, in which adenine could be involved directly (57). Competition experiments carried out with adenine analogs (61) have suggested that exogenous adenine might replace A38 that in the wild-type hairpin

ribozyme strongly contributes to the architecture of the active site and is supposed to take part in catalytic chemistry (63). Adenine-dependent hairpin ribozymes were further evolved to specifically cleave *Tpl2/Cot* mRNA encoding a serine threonine kinase that activates the mitogen-activated protein kinase (MAPK) pathway implicated in cell proliferation in breast cancer (58, 60).

SELF-PROCESSING HAIRPIN RIBOZYMES

As mentioned above, the characteristic cleavage/ligation properties of the hairpin ribozyme make it a superior tool for applications where RNA cleavage and ligation are needed to be combined in a scheme of subsequent reactions. One such application is twin ribozyme-mediated RNA repair (38) (see above). The twin ribozyme reaction can be also seen in the context of RNA world scenarios as a method for RNA recombination that would support the generation of sequence diversity. Similarly, at some point small RNA genomes would have transited to

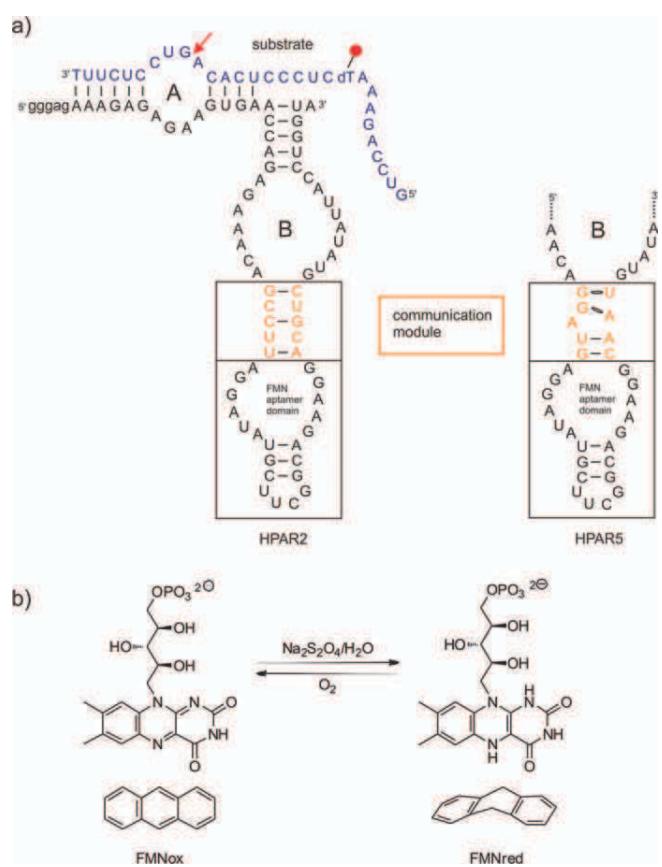


Figure 8. a) FMN dependent hairpin ribozymes HPAR2 and HPAR5 having distinct communication modules (52), b) Change of the FMN molecular geometry upon reduction of the isoalloxazin ring. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

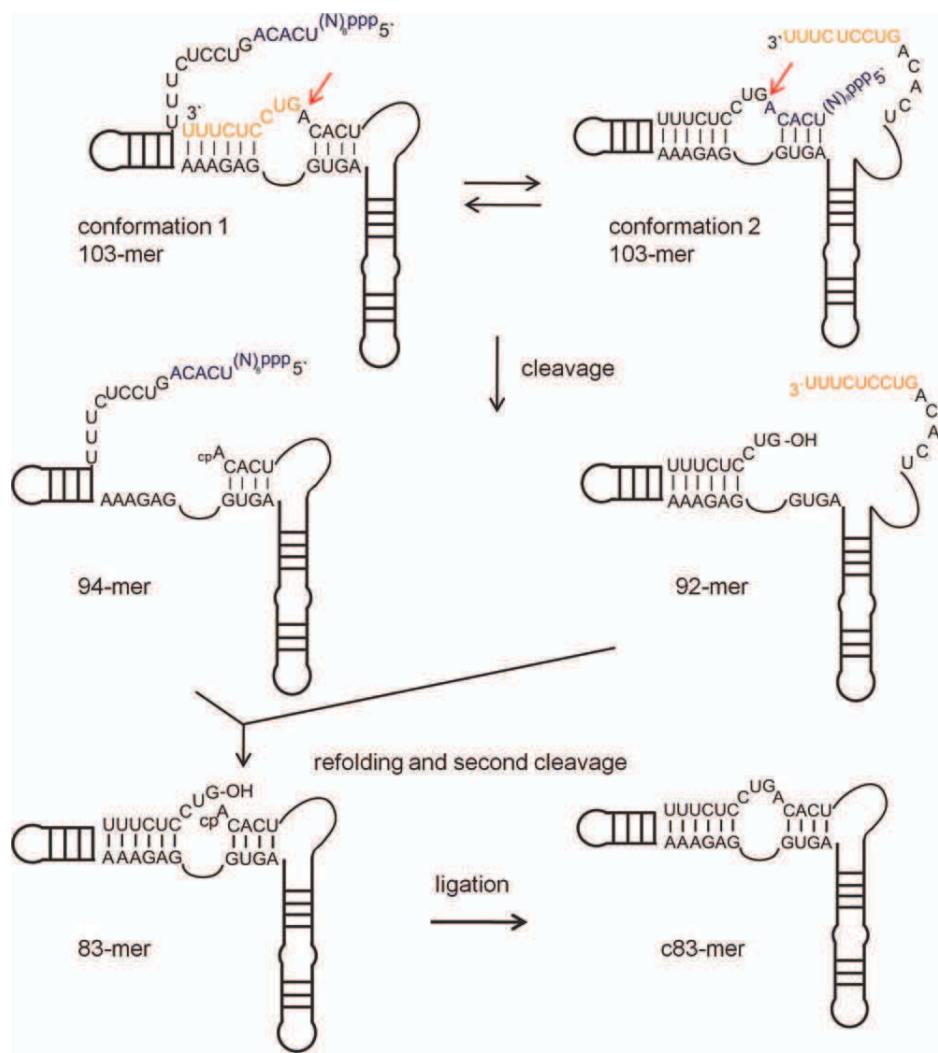


Figure 9. Schematic presentation of the reaction cascade of self-processing hairpin ribozymes (64). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

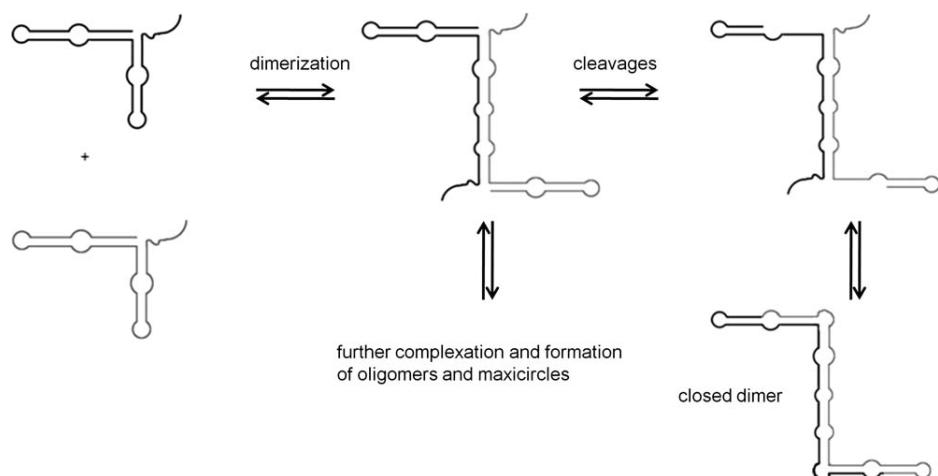


Figure 10. Organisation of self-processing hairpin ribozyme variants to higher order structures and intermolecular ligation (64).

larger RNA genomes requiring some kind of ligation reactions that join RNA pieces together. As described above, Vlassov et al. (33) have demonstrated ligation of a wide variety of RNA substrates by truncated and fragmented derivatives of the hairpin ribozyme at low temperatures. We have designed a more dynamic system that relies on hairpin ribozyme-mediated cleavage and ligation (64). Our initial attempt was to synthesize cyclic RNAs by controlled self-processing of hairpin ribozyme containing linear RNAs. However, in the course of experiments it turned out that the products resulting from RNA cleavage were not only prone to intramolecular but also to intermolecular ligation, leading to oligomeric versions of the parent RNA in addition to circles. The predicted reaction course is shown in Fig. 9.

RNA transcribed from a DNA template was programmed to fold in two distinct hairpin ribozyme derived cleavage-active conformations. Either the 3'-end (conformation 1) or the 5'-end (conformation 2) is removed first followed by the other. Thus, upon two cleavage events, a linear 83-mer with 5'-OH group and 2',3'-cyclic phosphate is produced. This 83-mer can fold back to form a ligation active structure generating a cyclic 83-mer by joining the two ends. As analyzed by electrophoresis through denaturing polyacrylamide gels, indeed all expected intermediates and products were observed in the reaction mixture. However, in addition also bands corresponding to strongly retarded species were observed in the gel. We have suggested that these bands represent oligomeric products resulting from intermolecular ligation, and/or cyclic dimers that would result from dimerization of the linear 83-mer followed by ligation at the two sites. This becomes plausible when one considers that (i) the hairpin ribozyme is known to form dimers with two active sites, and (ii) in particular the intermediates resulting from the first two cleavage events carry both a 5'-OH group and a 2',3'-cyclic phosphate, they can undergo second cleavage or ligation, and moreover, are able to self-organize into higher structures as depicted in Fig. 10. We are currently conducting new experiments and analytics to provide further proof of this initial result.

A related system of self-processing hairpin ribozyme was engineered by Diegelman and Kool for the purpose of production of circular RNAs and *trans*-cleaving hairpin ribozymes by rolling circle transcription (65). Circular DNA oligonucleotides encoding both the hairpin ribozyme and ribozyme cleavable sequences were used as template and transcribed *in vitro* by *E. coli* RNA polymerase. As a result of this reaction, long repeating RNAs were produced. The repetitive RNAs were found to undergo self-processing, yielding unit length linear and circular RNAs as key products. The linear products were able to cleave appropriate substrates in *trans*, while the circular RNAs were products of self-ligation. This result very well agrees with the observation made in our laboratory and described above: self-processed hairpin ribozymes can be catalytically active in *trans* despite the presence of self-binding domains.

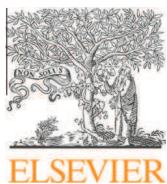
Taken together, all hairpin ribozyme variants described in this review demonstrate the strong relation between structure and function of this catalytic RNA. Structural extensions or changes have resulted in RNA catalysts with new properties. Thereby, the basic function, cleavage and ligation of RNA, was conserved. However, in a number of variants, activity became assessable. Furthermore, clever structural design has allowed construction of hairpin ribozyme derivatives with new functions. The many examples of ribozyme engineering demonstrate that we have understood these systems to a level that allows for adapting and subsequently using them for a specific purpose.

REFERENCES

- Guo, P. (2010) The emerging field of RNA nanotechnology. *Nat. Nanotechnol.* **5**, 833–842.
- Weigand, J. E., and Suess, B. (2009) Aptamers and riboswitches: perspectives in biotechnology. *Appl. Microbiol. Biotechnol.* **85**, 229–236.
- Link, K. H., and Breaker, R. R. (2009) Engineering ligand-responsive gene-control elements: lessons learned from natural riboswitches. *Gene Ther.* **16**, 1189–1201.
- Schubert, S., and Kurreck, J. (2004) Ribozyme- and deoxyribozyme-strategies for medical applications. *Curr. Drug Targets* **5**, 667–681.
- Shippy, R., Lockner, R., Farnsworth, M., and Hampel, A. (1999) The hairpin ribozyme. Discovery, mechanism, and development for gene therapy. *Mol. Biotechnol.* **12**, 117–129.
- Ferré-D'Amaré, A. R. (2004) The hairpin ribozyme. *Biopolymers* **73**, 71–78.
- Feldstein, P.A., Buzayan, J.M., and Bruening, G. (1989) Two sequences participating in the autolytic processing of satellite tobacco ringspot virus complementary RNA. *Gene* **82**, 53–61.
- Hampel, A., and Tritz, R. (1989) RNA catalytic properties of the minimum (-)sTRSV sequence. *Biochemistry* **28**, 4929–4933.
- Fedor, M. J. (1999) Tertiary structure stabilization promotes hairpin ribozyme ligation. *Biochemistry* **38**, 11040–11050.
- Rupert, P. B., and Ferré-D'Amaré, A. R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, 780–786.
- Walter, N. G., Hampel, K. J., Brown, K. M., and Burke, J. M. (1998) Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *EMBO J.* **17**, 2378–2391.
- Spitale, R. C., Volpini, R., Heller, M. G., Krucinska, J., Cristalli, G., et al. (2009) Identification of an imino group indispensable for cleavage by a small ribozyme. *J. Am. Chem. Soc.* **131**, 6093–6105.
- Suydam, I.T., Levandoaski, S. D., and Strobel, S. A. (2010) Catalytic importance of a protonated adenosine in the hairpin ribozyme active site. *Biochemistry* **49**, 3823–3732.
- Feng, Y., Leavitt, M., Tritz, R., Duarte, E., Kang, D., et al. (2000) Inhibition of CCR5-dependent HIV-1 infection by hairpin ribozyme gene therapy against CC-chemokine receptor 5. *Virology* **276**, 271–278.
- Drude, I., Strahl, A., Galla, D., Müller, O., and Müller, S. (2011) Design of hairpin ribozyme variants with improved activity for poorly processed substrates. *FEBS J.* **278**, 622–633.
- Pérez-Ruiz, M., Barosso-delJesus, A., and Berzal-Herranz, A. (1999) Specificity of the hairpin ribozyme. *J. Biol. Chem.* **274**, 29376–29380.
- Sargeul, B., Pecchia, D. B., and Burke, J.M. (1995) An improved version of the hairpin ribozyme functions as a ribonucleoprotein complex. *Biochemistry* **34**, 7739–7748.

18. Murchie, A. I. H., Thomson, J. B., Walter, F., and Lilley, D. M. J. (1998) Folding of the hairpin ribozyme in its natural conformation achieves close physical proximity of the loops. *Mol. Cell* **1**, 873–881.
19. Walter, N. G., Burke, J. M., and Millar, D. P. (1999) Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction. *Nat. Struct. Biol.* **6**, 544–549.
20. Walter, N. G., Chan, P. A., Hampel, K. J., Millar, D. P., and Burke, J. M. (2001) A base change in the catalytic core of the hairpin ribozyme perturbs function but not domain docking. *Biochemistry* **40**, 2580–2587.
21. Welz, R., Schmidt, C., and Müller, S. (2001) Spermine supports hairpin ribozyme variants to differing extents. *Biochem. Biophys. Res. Commun.* **283**, 648–654.
22. Komatsu, Y., Shirai, M., Yamashita, S., and Ohtsuka, E. (1997) Construction of hairpin ribozymes with a three-way junction. *Bioorg. Med. Chem.* **5**, 1063–1069.
23. Esteban, J. A., Walter, N. G., Kotzorek, G., Heckman, J. E., and Burke, J. M. (1998) Structural basis for heterogeneous kinetics: reengineering the hairpin ribozyme. *Proc. Natl. Acad. Sci. USA* **95**, 6091–6096.
24. Komatsu, Y., Kanzaki, I., Shirai, M., and Ohtsuka, E. (1997) A new type of hairpin ribozyme consisting of three domains. *Biochemistry* **36**, 9935–9940.
25. Earnshaw, D. J., and Gait, M. J. (1998) Hairpin ribozyme cleavage catalyzed by aminoglycoside antibiotics and the polyamine spermine in the absence of metal ions. *Nucleic Acids Res.* **26**, 5551–5561.
26. Stolze, K., Holmes, S. C., Earnshaw, D. J., Singh, M., Stetsenko, D., et al. (2001) Novel spermine-amino acid conjugates and basic tripeptides enhance cleavage of the hairpin ribozyme at low magnesium ion concentration. *Bioorg. Med. Chem. Lett.* **11**, 3007–3010.
27. Marsh, A. J., Williams, D. M., and Grasby, J. A. (2004) The synthesis and properties of oligoribonucleotide-spermine conjugates. *Org. Biomol. Chem.* **2**, 2103–2112.
28. Feldstein, P. A., and Bruening, G. (1993) Catalytically active geometry in the reversible circularization of ‘mini-monomer’ RNAs derived from the complementary strand of tobacco ringspot virus satellite RNA. *Nucleic Acids Res.* **21**, 1991–1998.
29. Komatsu, Y., Koizumi, M., Nakamura, H., and Ohtsuka, E. (1994) Loop-size variation to probe a bent structure of a hairpin ribozyme. *J. Am. Chem. Soc.* **116**, 3692–3696.
30. Butcher, S. E., and Burke, J. M. (1994) Structure-mapping of the hairpin ribozyme. Magnesium-dependent folding and evidence for tertiary interactions within the ribozyme-substrate complex. *J. Mol. Biol.* **244**, 52–56.
31. Butcher, S. E., Heckman, J. E., and Burke, J. M. (1995) Reconstitution of hairpin ribozyme activity following separation of functional domains. *J. Biol. Chem.* **270**, 29648–29651.
32. Shin, C., Choi, J. N., Song, S. I., Song, J. T., et al. (1996) The loop B domain is physically separable from the loop A domain in the hairpin ribozyme. *Nucleic Acids Res.* **24**, 2685–2689.
33. Vlassov, A. V., Johnston, B. H., Landweber, L. F., and Kazakov, S. A. (2004) Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world. *Nucleic Acids Res.* **32**, 2966–2974.
34. Komatsu, Y., Kanzaki, I., Koizumi, M., and Ohtsuka, E. (1995) Modification of primary structures of hairpin ribozymes for probing active conformations. *J. Mol. Biol.* **252**, 296–304.
35. Schmidt, C., Welz, R., and Müller, S. (2000) RNA double cleavage by a hairpin-derived twin ribozyme. *Nucleic Acids Res.* **28**, 886–894.
36. Ivanov, S. A., Vauléon, S., and Müller, S. (2005) Efficient RNA ligation by reverse-joined hairpin ribozymes and engineering of twin ribozymes consisting of conventional and reverse-joined hairpin ribozyme units. *FEBS J.* **272**, 4464–4474.
37. Ivanov, S. A., Volkov, E. M., Oretskaya, T. S., and Müller, S. (2004) Chemical synthesis of an artificially branched hairpin ribozyme variant with RNA cleavage activity. *Tetrahedron* **60**, 9273–9281.
38. Welz, R., Bossmann, K., Klug, C., Schmidt, C., Fritz, H.-J., et al. (2003) Site-directed alteration of RNA sequence mediated by an engineered twin ribozyme. *Angew. Chem. Int. Ed.* **42**, 2424–2427.
39. Vauléon, S., Ivanov, S. A., Gwiazda, S., and Müller S. (2005) Site-specific fluorescent and affinity labeling of RNA mediated by an engineered twin ribozyme. *ChemBioChem* **6**, 2158–2162.
40. Drude, I., Vauléon, S., and Müller, S. (2007) Twin ribozyme mediated removal of nucleotides from an internal RNA site. *Biochem. Biophys. Res. Commun.* **363**, 24–29.
41. Nesbitt, S. M., Erlacher, H. A., and Fedor, M. J. (1999) The internal equilibrium of the hairpin ribozyme: temperature, ion and pH effects. *J. Mol. Biol.* **286**, 1009–1024.
42. Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., et al. (1997) Constitutive transcriptional activation by ???-catenin-Tcf complex in APC – / – colon carcinoma. *Science* **275**, 1784–1787.
43. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., et al. (1997) Activation of β-catenin-Tcf signalling in colon cancer by mutations in β-catenin or APC. *Science* **275**, 1787–1790.
44. Drude, I., Bahlke, D., Strahl, A., Müller, O., and Müller, S. manuscript in preparation.
45. Link, K. H., and Breaker, R. R. (2009) Engineering ligand-responsive gene-control elements: lessons learned from natural riboswitches. *Gene Ther.* **16**, 1189–1291.
46. Silverman, S. K. (2003) Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA* **9**, 377–384.
47. Komatsu, Y., Nobuoka, K., Karino-Abe, N., Matsuda, A., and Ohtsuka, E. (2002) In vitro selection of hairpin ribozymes activated with short oligonucleotides. *Biochemistry* **41**, 9090–9098.
48. Vauléon, S., and Müller, S. (2003) External regulation of hairpin ribozyme activity by an oligonucleotide effector. *ChemBioChem* **4**, 220–224.
49. Najafi-Shoushtari, S. H., Mayer, G., and Famulok, M. (2004) Sensing complex regulatory networks by conformationally controlled hairpin ribozymes. *Nucleic Acids Res.* **32**, 3212–3219.
50. Najafi-Shoushtari, S. H., and Famulok, M. (2005) Competitive regulation of modular allosteric aptazymes by a small molecule and oligonucleotide effector. *RNA* **11**, 1514–1520.
51. Najafi-Shoushtari, S. H., and Famulok, M. (2007) DNA aptamer-mediated regulation of the hairpin ribozyme by human alpha-thrombin. *Blood Cells Mol. Dis.* **38**, 19–24.
52. Strohbach, D., Novak, N., and Müller S. (2006) Redox-active riboswitching: allosteric regulation of ribozyme activity by ligand shape control. *Angew. Chem. Int. Ed.* **45**, 2127–2129.
53. Strohbach, D., Turku, F., Schuhmann, W., and Müller, S. (2008) Electrochemically induced modulation of the catalytic activity of a reversible redoxsensitive riboswitch. *Electroanalysis* **20**, 935–940.
54. Fan, P., Suri, A. K., Fiala, R., Live, D., and Patel, D. J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.* **258**, 480–500.
55. Krellenberg, T., and Müller, S. manuscript in preparation.
56. Hall, B., Hesselberth, J. R., and Ellington, A. D. (2007) Computational selection of nucleic acid biosensors via a slip structure model. *Biosensor. Bioelectron.* **22**, 1939–1947.
57. Meli, M., Vergne, J., and Maurel, M. C. (2003) In vitro selection of adenine-dependent hairpin ribozymes. *J. Biol. Chem.* **278**, 9835–9842.
58. Li, Y. L., Torchet, C., Vergne, J., and Maurel, M. C. (2007) A cis and trans adenine-dependent hairpin ribozyme against Tpl-2 target. *Biochimie* **89**, 1257–1263.
59. Li, Y. L., Maurel, M. C., Ebel, C., Vergne, J., Pipich, V., et al. (2008) Self-association of adenine-dependent hairpin ribozymes. *Eur. Biophys. J.* **37**, 173–182.
60. Li, Y. L., Vergne, J., Torchet, C., and Maurel, M. C. (2009) In vitro selection of adenine-dependent ribozyme against Tpl2/Cot oncogene. *FEBS J.* **276**, 303–314.

61. Ztouti, M., Kaddour, H., Miralles, F., Simian, C., Vergne, J., et al. (2009) Adenine, a hairpin ribozyme cofactor--high-pressure and competition studies. *FEBS J.* **276**, 2574–2588.
62. Buck, J., Li, Y. L., Richter, C., Vergne, J., Maurel, M. C., et al. (2009) NMR spectroscopic characterization of the adenine-dependent hairpin ribozyme. *ChembioChem.* **10**, 2100–2110.
63. Salter, J., Krucinska, J., Alam, S., Grum-Tokars, V., and Wedekind, J. E. (2006) Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry* **45**, 686–700.
64. Pieper, S., Vauléon, S., and Müller S. (2007) RNA self-processing towards changed topology and sequence oligomerization. *Biol. Chem.* **388**, 743–746.
65. Diegelman, A. M., and Kool, E. T. (1998) Generation of circular RNAs and trans-cleaving catalytic RNAs by rolling transcription of circular DNA oligonucleotides encoding hairpin ribozymes. *Nucleic Acids Res.* **26**, 235–241.



RNA self-processing: Formation of cyclic species and concatemers from a small engineered RNA

Sonja Petkovic, Sabine Müller*

Ernst-Moritz-Arndt-Universität Greifswald, Institut für Biochemie, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany



ARTICLE INFO

Article history:

Received 10 April 2013

Revised 5 June 2013

Accepted 5 June 2013

Available online 22 June 2013

Edited by Michael Ibba

Keywords:

Cleavage

Ligation

Nucleic acid

Ribozyme

RNA

Self-processing

ABSTRACT

We have engineered a self-processing RNA, derived from the hairpin ribozyme that runs through a cascade of cleavage and ligation reactions thereby changing its topology. The first two cleavage events leave the resulting RNA with a 5'-OH group and a 2',3'-cyclic phosphate. Thus, upon refolding, intramolecular ligation delivers a cyclic species. In addition, we demonstrate formation of concatemers resulting from multiple intermolecular ligations. Our results demonstrate the potential of RNA for self-supported topology changes and support the suggestion of 2',3'-cyclic phosphates being suitable activated building blocks for reversible phosphodiester bond formation in the RNA world.

© 2013 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

Over the past two decades, RNA engineering has become a strong field of research. On one hand, synthetic RNAs have enabled sophisticated studies into structure and mechanism of RNAs with various cellular functions. On the other hand, understanding the mechanism of action of those RNAs has helped in the design of new functional entities composed of natural and/or modified ribonucleic acids. Research in the field of chemical biology for example has afforded a variety of ribozymes with useful activities, artificial riboswitches, and diagnostic tools [1–4]. Our contribution to the field have been a number of RNA systems derived from the hairpin ribozyme that allow for controlled cleavage and ligation of suitable RNA substrates [5–11]. In particular, we have developed twin ribozymes for RNA repair and site-specific labeling [5–8], and aptazymes that are controlled by the oxidation state of the allosteric activator flavine mononucleotide [9,10]. In the course of those studies, we also sought to develop a ribozyme that processes itself from a linear transcript to a cyclic RNA species [11]. Apart from the interest in RNA-self-processing, the development of a strategy for the cyclization of RNA to be used as nuclease resistant molecular tools was a strong motivation for this project. The primary RNA resulting from transcription was designed to undergo two cleavage

reactions followed by ligation as depicted here in Fig. 1 (103mer transcript → 92mer/94mer after first cleavage → 83mer after second cleavage → cyclic 83mer after intramolecular ligation). It is derived from the hairpin ribozyme [12–14], the cleavage/ligation characteristics of which are particularly suited to this purpose. The hairpin ribozyme cleavage reaction proceeds by nucleophilic attack of the 2'-OH on the neighboring phosphorous leading to a pentavalent intermediate (Fig. 2). From that, the 5'-oxygen of the neighboring nucleoside is released forming the final products, one carrying a 2',3'-cyclic phosphate, the other a 5'-OH group. Ligation proceeds via the same intermediate, thus requiring fragments with 5'-OH and 2',3'-cyclic phosphate as substrates (Fig. 2) [12–14]. The hairpin ribozyme is a rather rigid structure undergoing very little conformational change upon cleavage/ligation. Therefore, ligation being the faster reaction compared with cleavage, is favored if cleavage products/ligation substrates are tightly bound to the ribozyme. However, if the ribozyme structure is less stable, such that cleavage products/ligation substrates can easily dissociate, the cleavage reaction becomes dominant [15]. Based on these features, we have previously designed the 103mer transcript [11] to fold in the two conformations shown in Fig. 1a. Both structures should favor cleavage, because upon cleavage the short overhanging 5'-, and 3'-terminal sequences can easily dissociate. According to hairpin ribozyme characteristics, cleavage of the 5'-terminal sequence leaves a 5'-OH group and cleavage of the 3'-terminal sequence leaves a 2',3'-cyclic phosphate at the remaining 83mer

* Corresponding author. Fax: +49 03834 864471.

E-mail address: smueller@uni-greifswald.de (S. Müller).

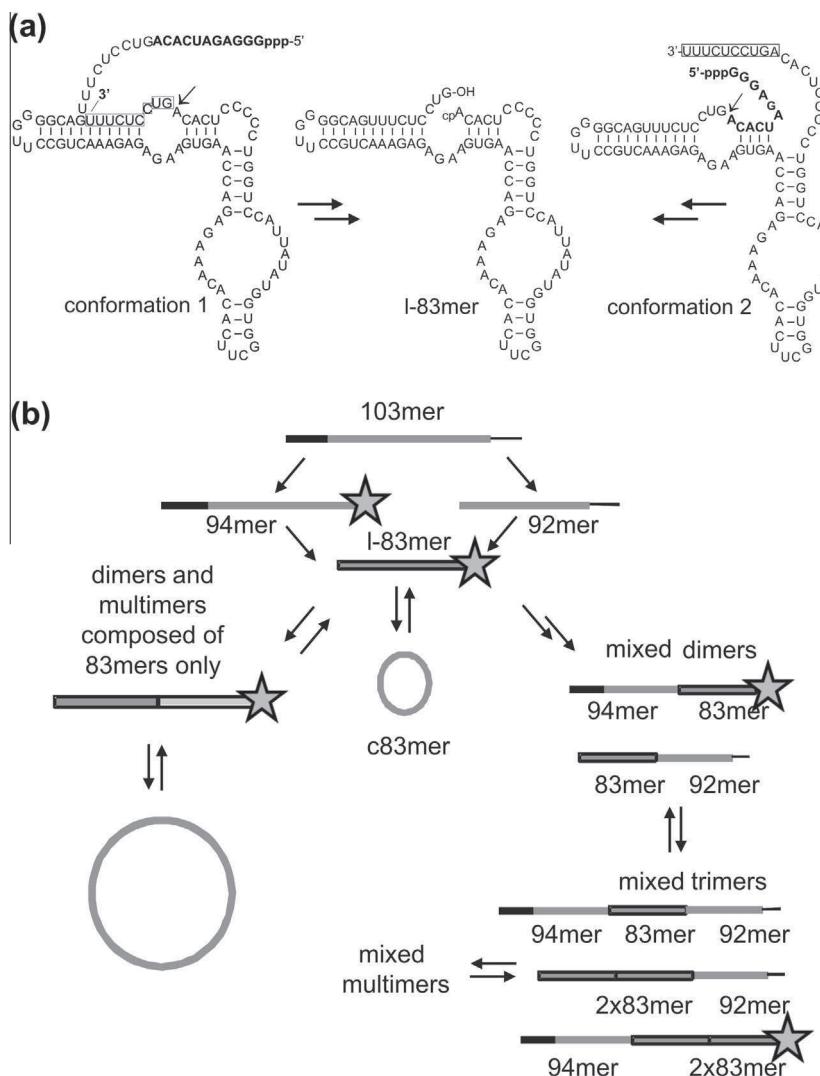


Fig. 1. Self-processing RNA CRZ-2 derived from the hairpin ribozyme. (a) Sequence of CRZ-2 folded in two alternative conformations (left and right structure). Upon two subsequent cleavage events, both RNA structures are converted in the shortened sequence shown in the middle. Cleavage sites are marked with an arrow. (b) Reaction path of self-processing. The original 103mer RNA derived from the hairpin ribozyme cleaves itself to remove the 5'- and 3'-ends. Resulting from a typical hairpin ribozyme reaction, the produced 83mer has a 5'-OH group and a 3'-terminal 2',3'-cyclic phosphate. Due to the intrinsic hairpin ribozyme activity, it can form cyclic species by intramolecular ligation or concatemers by self-association and intermolecular ligation. Cyclic phosphates are marked with an asterisk.

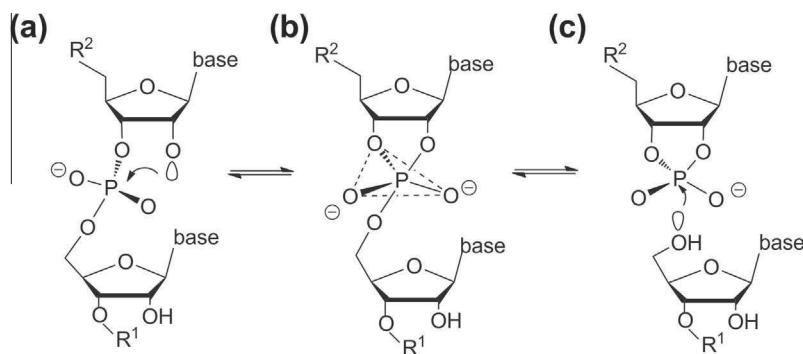


Fig. 2. Mechanism of the hairpin ribozyme catalyzed cleavage and ligation of RNA. Both reactions proceed via the same pentavalent intermediate, generating characteristic fragments with 2',3'-cyclic phosphate and 5'-OH after cleavage, and requiring these functionalities for ligation. R^1 , R^2 = RNA chain.

(Fig. 1). Thus, the produced 83mer still possesses all elements required for hairpin ribozyme activity, and now should favor ligation due to the intramolecular arrangement of the ligation site hampering dissociation. The final ligation would lead to the expected cyc-

lic RNA. To our surprise, upon reaction analysis, we had detected a variety of products, very likely resulting from intermolecular ligation in addition to the predicted intramolecular cyclization [11]. This is a very interesting result, because it shows the potential of

RNA for self-organization and polymerization. Intermolecular ligation apparently is in strong competition with intramolecular reaction, and thus might be considered as a model of self-supported generation of longer RNAs from smaller fragments in the RNA world. We decided to further study the potential of this RNA for concatemerization versus cyclization in order to possibly define conditions that allow controlling the tendency towards the one or the other activity. As a first step towards this goal, we here present an analysis of the formed products and provide support of their cyclic or linear nature.

2. Materials and methods

2.1. General information

All aqueous solutions were made with demineralised water which was further purified with a Barnstedt Nanopure system and autoclaved. Deoxynucleotide triphosphates (dNTPs), nucleotide triphosphates (NTPs), Klenow buffer, DNase I, Klenow fragment exo-, *Sma*I, RiboLock™, RiboRuler low range RNA ladder and polynucleotide kinase (PNK) were purchased from Fermentas Company; RNA Ligase (T4 RnL2) and Vent® polymerase were obtained from New England Biolabs. Dried phosphodiesterase I (PDE I) from *Crotalus atrox* was purchased from Sigma-Aldrich®. Agar, X-Gal and IPTG were purchased from Roth. DNA primers were provided by Biomers. All reagents were of analytical grade and filtered before use.

Measures for glass plates of polyacrylamide gel electrophoresis were $200 \times 150 \times 1.5$ mm³. All UV spectra were recorded on a NanoDrop ND 1000 spectrophotometer. Product band intensities were quantified using Gene ImageIR 4.05.

2.2. RNA synthesis

Klenow primers (P1: 5'-AAA GAG GAC TGT GAG GGG GAC CAG GTA ATA TAC CAC CGA AGT GTG *TTT* CTC TGG *TTC* ACT TCT C-3' P2: 5'-TAA TAC GAC TCA CTA TAG GGA GAT CAC AGT CCT CTT TGA CGG GGT TCC GTC AAA GAG AGA AGT GAA CCA GAG AAA CA-3') with 21 bp overlap shown in italics were used in the Klenow reaction with Klenow exo-polymerase from Fermentas company following the manufacturer's protocol. The obtained dsDNA was subjected to native agarose gel electrophoresis (1.5%, EtBr stained), product containing bands were cleaved off, and DNA was purified utilizing the QIAquick gel extraction kit (Qiagen).

RNAs were synthesized by in vitro transcription of double stranded DNA templates (1 μM) with T7 RNA polymerase (35–70 ng/μl, self-made) in the presence of the four ribonucleoside triphosphates (2 mM) and 1 U/μl RiboLock™ and 1× HEPES buffer in a total reaction volume of 50 μl for 3 h at 37 °C. The DNA template was digested using DNasel following the manufacturer's protocol. Final purification was achieved by electrophoresis through 15% denaturing polyacrylamide gels, followed by elution of the product with sodium acetate (0.1–0.3 M, pH 7 at 10 °C) and precipitation with ethanol overnight at –20 °C.

2.3. Synthesis of the inactive linear and cyclic variant of the 83mer

The inactive 83mer (in-l-83) was transcribed from the dsDNA template: 5'-TAA TAC GAC TCA CTA TA **GGG AGA** GTG CTC CCA AAG AGG **ACT** GTG AAC CAG AGA AAG ACA CTT CGG TGG TAT ATT ACC TGG TCC CCC TCA CAG TCC TCT TT-3' (Sequence of the sense strand is given; T7 promoter sequence in italics, sequence changes in comparison to the linear 83mer resulting from self-processing of the 103mer transcript are indicated in bold). For in vitro transcription, GMP was added to the NTP mix at a 4.8:1 ratio of

GMP/GTP. Transcription was carried out following the protocol of Harris and Christian for incorporation of guanosine monophosphorothioate [16].

For cyclization of in-l-83, enzymatic ligation was conducted for 4 h at native conditions using T4 RnL2 in the supplied buffer (composed of 50 mM Tris-HCl, 2 mM MgCl₂, 1 mM DTT, 400 μM ATP, pH 7.5) in a total reaction volume of 20 μl following the suppliers protocol.

2.4. Ribozyme self-processing reaction

RNA-transcripts or isolated fragments (50 pmol) were taken up in Tris-HCl buffer (10 mM, pH 7.5). After denaturation for one minute at 90 °C, RNA folding for 15 min at room temperature followed. To initiate the cleavage reaction, MgCl₂ hexahydrate to a final concentration of 10 mM was added and cleavage took place within 2 h at 37 °C. For preferential ligation, Mg²⁺ concentration was increased up to 50 mM, and reaction proceeded for 2, 4 or 24 h at 37 °C. The final concentration of RNA in the self-processing reaction was 2 μM. Reaction was stopped by ethanol precipitation over night or 1:1 addition of stop mix composed of urea (7 M) and EDTA (50 mM) for direct PAGE analysis.

2.5. Two-dimensional denaturing gel electrophoresis

RNA (50 pmol) was analyzed using two-dimensional PAGE (first dimension, 12% polyacrylamide; second dimension, 16% polyacrylamide) under denaturing conditions (7 M urea). After the second electrophoresis, the ethidium bromide-stained gel was visualized using Chemi-Smart 2000 WL/LC 26M.

2.6. Sequencing analysis

RNA was reverse transcribed by SuperScript™ III Reverse Transcriptase (Invitrogen) with RT-primer, 5'-ATA GGA TCC AAA GAG GAC TGT GAG GGG G-3' according to the manufacturer's protocol. Non-purified single stranded cDNA was amplified with Taq (0.025 U/μl, self-made) and Vent® (0.01 U/μl) DNA polymerase with the RT-primer, reverse primer, 5'-ATA TCT AGA GAC GGG GTT CCG TCA AAG-3', dNTPs (1 μM) and 1× Thermo Pol PCR buffer 4S in a total reaction volume of 100 μl. Amplified double stranded DNA was phosphorylated using 0.5 μg DNA, supplied reaction buffer A (with final concentrations: 50 mM Tris-HCl pH 7.6 10 mM MgCl₂, 5 mM DTT and 0.1 mM spermidine) 1 mM ATP and 0.5 U/μl PNK in a total volume of 20 μl at 37 °C for 30 min, followed by heat inactivation of the PNK applying 70 °C for 10 min. Phosphorylated DNA was purified using QIAgen PCR purification kit and ligated into the *Sma*I restriction site of pUC18 (Stratagene). Ligation was conducted using a relation of 1:3 for vector: insert. 5 ng/μl plasmid and 15 ng/μl insert were added to 20 U of T4 DNA ligase in the supplied reaction buffer (composed of 50 mM Tris-HCl 10 mM MgCl₂, 1 mM ATP 10 mM DTT) in a total volume of 20 μl. Ligation proceeded for 16 h, starting at 18 °C for 3 h, followed by decreasing the temperature by 2 grad every 3 h. After 9 h, temperature was decreased to 10 °C and hold for 2 h and finally to 4 °C and hold for another 2 h. Then, the mixture was treated at 70 °C for 10 min for heat inactivation. The resulting plasmid was transformed into *E. coli* TG1 competent cells (cc) using a heat shock transformation protocol. Two hundred and fifty microliter of ccTG1 were carefully mixed with 20 μl of ligation mix and left on ice for 30 min, followed by immediate incubation at 42 °C for 45 s, and then putting the cells back on ice for additional 5 min. Growing was reinitiated at 37 °C in a shaking incubator at 200 rpm for 1 h. Agar plates including Ampicillin with a final concentration of 100 μg/ml were treated with X-Gal (40 mg/ml – 40 μl per plate) and IPTG (0.25 M -- 30 μl per plate). Colonies of interest were

identified via blue-white screening, and the DNA was purified and sequenced by GATC Biotech AG.

2.7. Digestion with phosphodiesterase

The isolated linear 83mer (2.25 μ M) in 150 μ l ligation buffer (10 mM Tris-HCl pH 7.5, 50 mM MgCl₂) was incubated at 37 °C for two hours. Lyophilized PDE I was solved in water and 2 μ l were added to the RNA mixture to a final concentration of 0.7 μ g/ μ l. The same amount of water was added to a control reaction mixture. The first aliquot (25 μ l) was taken immediately after mixing the reagents. It was added to an equal volume of stop mix (7 M urea and 50 mM EDTA) and frozen in liquid nitrogen (time zero). Further aliquots were taken every two minutes and treated as described above. Samples were analyzed by PAGE (denaturing, 15%) after heat denaturation at 90 °C for two minutes.

2.8. Gene bank indices: EMBL Nucleotide Sequence Database

Trimer: Accession#: HF565114

Exemplary dimer: Accession#: HF565115

Exemplary dimer: Accession#: HF565116

3. Results and discussion

We have prepared the 103mer transcript (now called CRZ-2) and followed its processing reaction by polyacrylamide gel electrophoresis (Fig. 3). Processing is observed already during *in vitro* transcription as the characteristic product bands occur (Fig. 3a, lane 1). This is not surprising, since transcription proceeds at 37 °C, at pH 7.5, and in the presence of Mg²⁺ and other cations, conditions that also support hairpin ribozyme activity [14]. From the *in vitro* transcription mixture, the full lengths 103mer transcript as well as the presumed linear 94-, 92- and 83mer fragments were isolated and incubated at typical hairpin ribozyme cleavage conditions (Fig. 3b). As expected, the 103mer was cleaved again to the 92/94mer and 83mer, and the 92/94mer was further cleaved to the 83mer, as shown by the corresponding bands in lane 1 and 2. However, in addition to the incubated fragments, a number of bands with strongly retarded migration were observed (Fig. 3b, bands **a–e**). As already mentioned above, we hypothesized that these bands represent oligomeric products of the linear 83mer,

92mer and 94mer. Oligomerization would require association of the linear fragments to form active complexes, which is well possible due to sequence complementarities. Furthermore, dimerization of the hairpin ribozyme was described previously [17], and thus it is well possible that self-association followed by intermolecular reactions plays also a role in the system described here. This in turn would explain the production of concatemers by intermolecular ligation. First support for this suggestion, in our system was obtained from sequence analysis. The slowly migrating products in the range of 200–400 bp (Fig. 3a) were isolated and pooled, reverse transcribed and amplified for cloning and sequencing (in Supplementary Fig. S1). Sequence analysis showed the presence of dimers and trimers (for sequence codes see Section 2). However, no concatemers longer than trimers were found. A most likely explanation for this observation is that due to the presence of Mg²⁺ in the isolation and reverse transcription buffers, longer concatemers were re-cleaved to trimers, dimers and monomers as shown by the pattern of cDNA obtained from reverse transcription of isolated species (Supplementary Fig. S1). Furthermore, RT-PCR amplification of a concatemer is quite challenging, because the PCR primers will not only bind to the 3'-end of the template but also to internal sites at each repeating sequence unit of the concatemer. This in turn will produce a majority of monomer units instead of concatemers after amplification. From the mixture of the obtained cDNAs, after native polyacrylamide gel electrophoresis only the upper bands corresponding to the designated longer concatemers were isolated (in Supplementary Fig. S1), cloned and sequenced. In addition to full length double stranded concatemers, these upper bands in a native gel may also represent nicked duplexes containing several monomer strands bound to the corresponding concatemeric strand, such that for cloning and sequence analysis monomer units are over-represented. Indeed, out of 34 sequenced samples, 20 monomers, 13 dimers and only one trimer were found. Thus, there is no direct evidence for the existence of long concatemers. However, due to the clear detection of di- and trimers, the system's general capability of self-processing involving concatemerization has been shown.

All following analyzes were done exclusively with the isolated linear 83mer. It cannot undergo further cleavage; however, ligation can take place leading to a cyclic monomer and/or linear and cyclic oligomers, due to the presence of both, the 5'-OH group and the 3'-terminal cyclic phosphate as described above (Fig. 1). First, we analyzed the products being represented by bands **b–e** (Fig. 3) resulting from incubation of the isolated 83mer (band **a**). As a first attempt to differentiate between the linear and cyclic 83mer, we prepared an inactive variant of the linear 83mer by introduction of a number of sequence changes (see Section 2), and part of that was enzymatically circularized. In order to generate the 5'-phosphate required for circularization by enzymatic ligation, *in vitro* transcription was carried out in the presence of guanosine monophosphate as transcription initiator, thus delivering RNAs with 5'-monophosphate instead of the 5'-triphosphate (see Section 2). Circularization was performed with T4 RNA ligase II. The linear species (in-l-83) runs faster through a 15% denaturing polyacrylamide gel than the circular species (in-c-83) (Fig. 4). Comparison of this pattern with the bands resulting from the ribozyme reaction confirms the identity of band **a** being the linear 83mer. However, it has to be noted that the cyclic inactive 83mer (lane 4) has no corresponding band in lane 3. Apparently, there is no cyclic 83mer formed upon the ribozyme reaction, or it shows a different migration behavior in the gel compared with the inactive cyclic 83mer. The latter assumption may apply, because the sequence changes made for inhibition of ribozyme activity are likely to influence the structure of the cyclic species, and this in turn would result in distinct retention of the two cyclic species, in spite of both having the same size.

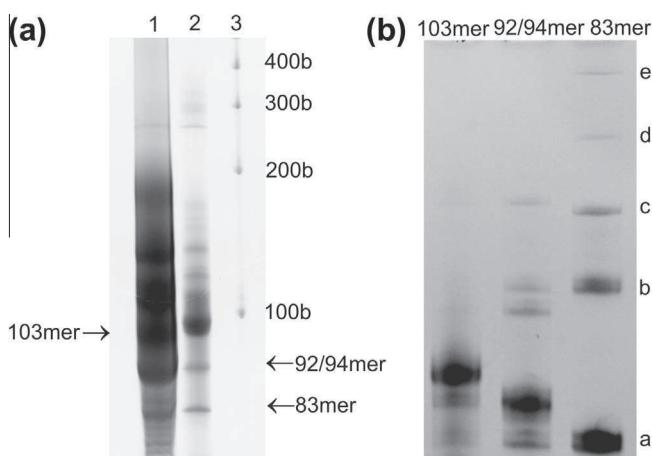


Fig. 3. In vitro transcription and self-processing of CRZ-2. (a) Fifteen percentage denaturing polyacrylamide gel, lane 1: *in vitro* transcription, lane 2: incubation of the isolated full length transcript in reaction buffer (10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 37 °C, 2 h (cleavage) and 50 mM MgCl₂, 37 °C 2 h (ligation)), lane 3: RNA ladder; (b) Re-incubation of isolated 103mer, 92/94mer and 83mer in cleavage buffer at 37 °C for 2 h. Individual species resulting from reaction of the 83mer were labeled **a–e**.

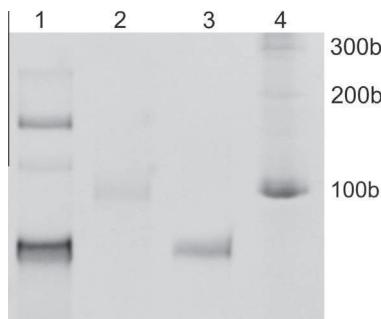


Fig. 4. Fifteen percentage polyacrylamide gel analysis of a separately synthesized inactive cyclic (lane 2) and linear (lane 3) 83mer in comparison to the ribozyme-derived active 83mer (lane 1). Lane 4: RNA ladder.

The most obvious way to differentiate between linear and cyclic nucleic acids is exonucleolytic digestion. However, analysis of isolated species is not possible, since the cyclic RNA upon isolation from the reaction mixture would undergo immediate re-cleavage to its linear counterpart. The linear RNA would be removed by exonuclease digestion, thus evoking further re-cleavage until the cyclic species is consumed and all RNA is digested. Therefore, we added phosphodiesterase I (PDE I) directly to the reaction mixture. This procedure cannot prevent re-cleavage of cyclic species as soon as linear species in the mixture are digested; however, the decay reaction can be better temporarily followed. To this end, after addition of the exonuclease, at time intervals indicated in Fig. 5, aliquots were taken out of the reaction mixture, and reaction was stopped and analyzed by gel electrophoresis. Bands **b**, **d** and **e** quickly disappeared, whereas band **c** was detectable over 4 min, and only then slowly disappeared from the gel. This implies that band **c** represents a cyclic species that first has to undergo re-cleavage before it can be digested, whereas bands **b**, **d** and **e** being digested immediately, correspond to linear RNAs. Band **a**, representing the linear 83mer, as was confirmed by the previous experiment (Fig. 4), is also still detectable after 4 min of PDE treatment, very likely because of its high excess in the reaction mixture.

In order to further confirm the result of this analysis, a two-dimensional PAGE experiment [18–20] was conducted. It was shown previously that migration of linear and circular nucleic acid species is distinctly dependent on the gel pore size [21]. Therefore, electrophoresis was run in the first dimension using 12% polyacrylamide, followed by 16% polyacrylamide in the second dimension. As a result, all linear species should form a diagonal, whereas circular RNAs are expected to be located beyond. In order to better fit the diagonal to linear species, an RNA size standard, being exclu-

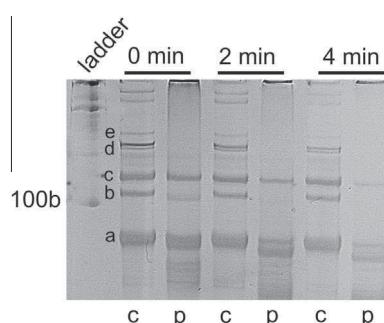


Fig. 5. Analysis of the exonucleolytic digestion of the RNA mixture obtained from incubation of the linear 83mer in ligation buffer (50 mM MgCl₂, 37 °C, 2 h). Aliquots were taken at the indicated time intervals and subjected onto a 15% denaturing polyacrylamide gel. Left lane: RNA ladder, lanes marked with c display control reactions including water instead of PDE, lanes marked with p display RNA reaction mixtures including PDE.

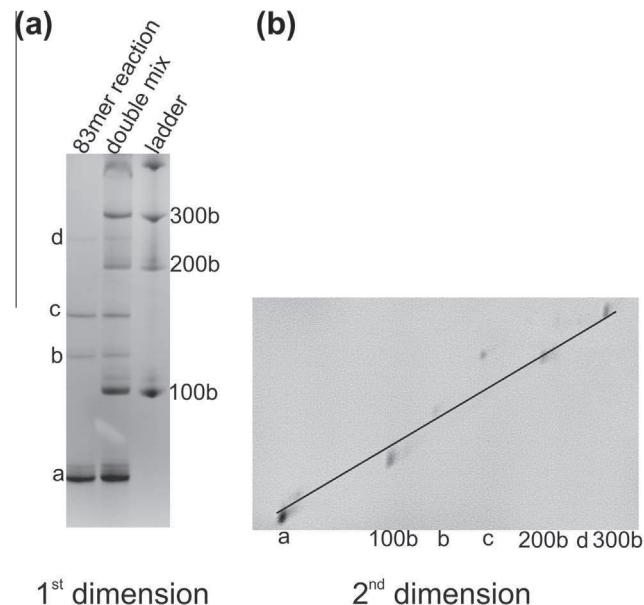


Fig. 6. 2D polyacrylamide gel electrophoretic analysis of self-processing of the linear 83mer using 12% (first dimension a) and 16% (second dimension, b) denaturing polyacrylamide gels. The self-processing reaction was left to proceed for 4 h in ligation buffer at 37 °C. For better visualization, the reaction mixture of the 83mer self-processing reaction was mixed with the RNA ladder prior to subjecting onto the gel. Individual species were labeled as **a–e** according to Fig. 3.

sively composed of linear RNA sequences, was added to the reaction mixture (double mix). As shown in Fig. 6, apart from species **c**, which clearly stands out, all other spots form a diagonal with the fragments of the RNA size standard, thus further supporting the cyclic nature of **c**.

Lastly, we calculated the relative amount of species occurring at different concentrations of the starting material (in Supplementary Fig. S2). Intermolecular reactions, hence the formation of oligomers, would profit from higher concentrations of the linear 83mer, whereas intramolecular cyclization is independent of concentration. Reaction was carried out at 0.4, 2 and 10 μM concentration of the linear 83mer (Supplementary Fig. S2). The intensity of individual bands was measured and the relative percentage amount within each lane was calculated (Table 1). In general, reaction is enhanced at higher concentration being mirrored in the higher consumption of the linear 83mer (species **a**), and thus indicating a considerable part of intermolecular reaction. In addition, it is seen that the relative amount of species **b** increases from the 0.4 μM reaction to the 10 μM reaction. On the opposite, the relative amount of species **c** is kept constant within error. Even though, this quantification has a rather high error rate, it provides further support to the interpretation made above that species **b** is a linear oligomer of the starting 83mer (presumably the dimer), while species

Table 1
Topology analysis. Calculation of the relative percentage of individual species in the reaction mixtures at 0.4, 2 and 10 μM starting concentration of the linear 83mer.

Band ^a	0.4 μM ^b	2 μM ^b	10 μM ^b	Suggested topology
e	1.2	2.0	4.5	—
d	8.2	7.3	7.0	—
c	8.1	6.7	9.5	Circular
b	4.2	13.5	16.2	Linear
a	78.3	70.5	62.8	Linear (83mer educt)

^a Bands **a–e** according to Fig. 4.

^b Concentration of the linear 83mer as starting material. Individual numbers represent the relative percentage of species **a–e** in the respective reaction mixture.

c is a circularized product (presumably the circularized monomer). We cannot reliably determine the nature of species **d** and **e**. There is no information from the 2D experiment on the nature of these two species, since the concentration of **d** and **e** was too low to be reliably detected after electrophoresis in the second dimension. According to the data in Table 1, **d** could be a cyclic species, whereas **e** is rather linear. From the PDE digestion experiment discussed above, one would however suggest that both **d** and **e** are linear RNAs, since both disappeared immediately after PDE treatment.

In conclusion, we have demonstrated the high potential of a small engineered RNA derived from the hairpin ribozyme for self-processing forming circles and concatemers of different size. This once more underlines the power of RNA for self-supported topology changes and sequence oligomerization. Furthermore, our results support the previous suggestion of 2',3'-cyclic phosphates being suitable activated building blocks for reversible phosphodiester bond formation [22], even more so, if catalytic RNA structures support the reaction. Our results are in agreement with previous reports emphasizing the importance of cyclic phosphates in the RNA world by demonstrating the formation of ribonucleotides with 2',3'-cyclic phosphate from a small collection of pre-biotic stock molecules [23,24].

Furthermore, the designed system demonstrates a potential strategy for easy preparation of cyclic RNA molecules. Due to their higher nuclease stability compared with linear RNAs, cyclic RNA molecules are advantageous for intracellular application of aptamers, ribozymes or other functional RNAs in molecular medicine and diagnostics [25–27]. Although, the reaction cascade described here does not deliver a stable cyclic RNA. Due to the intrinsic activity of the hairpin ribozyme, upon isolation the formed cycle is re-cleaved to its linear counterpart until equilibrium is reached. Preventing re-cleavage of the cyclic RNA requires switching hairpin ribozyme activity in a strictly controlled way. This is technically challenging, but might be manageable by engineering a self-processing RNA variant that is controlled by an external effector. We have shown previously [29] that activity of an inactive hairpin ribozyme variant can be switched on by an external oligonucleotide cofactor compensating for missing sequence elements required for activity. Implementing this principle in the system described here would have the advantage that upon removal of the external cofactor, activity is switched off, thus rendering the cyclized RNA stable. Furthermore, even though the hairpin ribozyme requires a number of conserved nucleotides for activity, large parts of the sequence can be varied as desired [28]. Thus, stable cyclic versions of specific aptamers or other RNAs of interest may be prepared by the strategy shown here combined with external activity control. In addition, the competitive oligomerization may be repressed by computer aided sequence optimization. Work along this line is in progress.

Acknowledgement

This work was supported by a DFG grant to S. Müller (MU 1396/11-1). The authors wish to thank Ivo Hofacker, Christoph Flamm and Stefan Badelt for fruitful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.febslet.2013.06.013>.

References

- [1] Wieland, M., Ausländer, D. and Fussenegger, M. (2012) Engineering of ribozyme-based riboswitches for mammalian cells. *Methods* 56, 351–357.
- [2] Liang, J.C., Bloom, R.J. and Smolke, C.D. (2011) Engineering biological systems with synthetic RNA molecules. *Mol. Cell* 43, 915–926.
- [3] Silvermann, S.K. (2008) Catalytic DNA (deoxyribozymes) for synthetic applications-current abilities and future prospects. *Chem. Commun.* 14, 3467–3485.
- [4] Wieland, M. and Hartig, J.S. (2008) Artificial riboswitches: synthetic mRNA-based regulators of gene expression. *ChemBioChem* 9, 1873–1878.
- [5] Welz, R., Bossmann, K., Klug, C., Schmidt, C., Fritz, H.-J. and Müller, S. (2003) Site-directed alteration of RNA sequence mediated by an engineered twin ribozyme. *Angew. Chem., Int. Ed.* 42, 2424–2427.
- [6] Ivanov, S.A., Vauléon, S. and Müller, S. (2005) Efficient RNA ligation by reverse-joined hairpin ribozymes and engineering of twin ribozymes consisting of conventional and reverse-joined hairpin ribozyme units. *FEBS J.* 272, 4464–4474.
- [7] Vauléon, S., Ivanov, S.A., Gwiazda, S. and Müller, S. (2005) Site-specific fluorescent and affinity labelling of RNA by using a small engineered twin ribozyme. *ChemBioChem* 6, 2158–2160.
- [8] Drude, I., Vauléon, S. and Müller, S. (2007) Twin ribozyme mediated removal of nucleotides from an internal RNA site. *Biochem. Biophys. Res. Commun.* 363, 24–29.
- [9] Strohbach, D., Nowack, N. and Müller, S. (2006) Redox-active riboswitching: allosteric regulation of ribozyme activity by ligand-shape control. *Angew. Chem., Int. Ed.* 45, 2127–2129.
- [10] Strohbach, D., Turcu, F., Schuhmann, W. and Müller, S. (2008) Electrochemically induced modulation of the catalytic activity of a reversible redoxsensitive riboswitch. *Electroanalysis* 20, 935–940.
- [11] Pieper, S., Vauléon, S. and Müller, S. (2007) RNA self-processing towards changed topology and sequence oligomerization. *Biol. Chem.* 388, 743–746.
- [12] Feldstein, P.A., Buzayan, J.M. and Bruening, G. (1989) Two sequences participating in the autolytic processing of satellite tobacco ringspot virus complementary RNA. *Gene* 82, 53–61.
- [13] Hampel, A. and Tritz, R. (1989) RNA catalytic properties of the minimum (-)sTRSV sequence. *Biochemistry* 28, 4929–4933.
- [14] Liley, D.M. (2011) Catalysis by the nucleolytic ribozymes. *Biochem. Soc. Trans.* 39, 641–646.
- [15] Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science* 296, 1473–1476.
- [16] Harris, M. and Christian, E. (1999) Use of circular permutation and end modification to position photoaffinity probes for analysis of RNA structure. *Methods* 18, 51–59.
- [17] Butcher, S.E. and Burke, J.M. (1994) A photo-cross-linkable tertiary structure motif found in functionally distinct RNA molecules is essential for catalytic function of the hairpin ribozyme. *Biochemistry* 33, 992–999.
- [18] Gunnarsson, G.H., Thormar, H.G., Gudmundsson, B., Akesson, L. and Johnsson, J.J. (2004) Two-dimensional conformation-dependent electrophoresis (2D-CDE) to separate DNA fragments containing unmatched bulge from complex DNA samples. *Nucleic Acids Res.* 32, e23.
- [19] Gunnarsson, G.H., Gudmundsson, B., Thormar, H.G., Alfredsson, A. and Johnsson, J.J. (2006) Two-dimensional strandness-dependent electrophoresis: a method to characterize single-stranded DNA, double-stranded DNA, and RNA-DNA hybrids in complex samples. *Anal. Biochem.* 350, 120–127.
- [20] Gunnarsson, G.H., Gudmundsson, B., Thormar, H.G., Alfredsson, A. and Johnsson, J.J. (2006) Two-dimensional strandness-dependent electrophoresis. *Nat. Protoc.* 1, 3011–3018.
- [21] Umekage, S. and Kikuchi, Y. (2009) In vitro and in vivo production and purification of circular RNA aptamer. *J. Biotechnol.* 139, 265–272.
- [22] Lutay, A.V., Zenkova, M.A. and Vlassov, V.V. (2007) Nonenzymatic recombination of RNA: possible mechanism for the formation of novel sequences. *Chem. Biodiversity* 4, 762–767.
- [23] Powner, M.W., Gestland, B. and Sutherland, J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459, 239–242.
- [24] Powner, M.W., Sutherland, J.D. and Szostak, J.W. (2010) Chemoselective multicomponent one-pot assembly of purine precursors in water. *J. Am. Chem. Soc.* 132, 16677–16688.
- [25] Sundaram, P., Kurniawan, H., Byrne, M.E. and Wower, J. (2012) Therapeutic RNA aptamers in clinical trials. *Eur. J. Pharm. Sci.* 48, 259–271.
- [26] Zhou, J., Bobbin, M.L., Burnett, J.C. and Rossi, J.J. (2012) Current progress of RNA aptamer-based therapeutics. *Front. Genet.* 3, 234.
- [27] Thiyyanathan, V. and Gorenstein, D.G. (2012) Aptamers and the next generation of diagnostic reagents. *Proteomics Clin. Appl.*, <http://dx.doi.org/10.1002/prca.201200042>.
- [28] Drude, I., Strahl, A., Galla, D., Müller, O. and Müller, S. (2011) Design of hairpin ribozyme variants with improved activity for poorly processed substrates. *FEBS J.* 278, 622–633.
- [29] Vauléon, S. and Müller, S. (2003) External regulation of hairpin ribozyme activity by an oligonucleotide effector. *ChemBioChem* 4, 220–224.

Supplementary Data

RNA self-processing: formation of cyclic species and concatemers from a small engineered RNA

Sonja Petkovic and Sabine Müller

Ernst-Moritz-Arndt-Universität Greifswald

Institut für Biochemie

Felix-Hausdorff-Str. 4

17487 Greifswald, Germany

Fax: 49 3834 864471

Tel: 49 3834 8622843

E-mail: smueller@uni-greifswald.de

1. Preparation of samples for sequence analysis

For sequence analysis, upper bands indicated as **c**, **d** and **e** in Figure S1 (left) were cleaved from the gel, RNAs were eluted using 0.3 M sodium acetate pH=7.0 with shaking at 600 rpm for 2 h at 10 °C. The process was repeated three times with fresh solution of sodium acetate. The RNA was precipitated from ethanol, pooled and reverse transcribed as described in section 1.5. After RT-PCR of the pooled RNAs, the gel (Fig. S1, right) showed again bands corresponding to the previously observed shorter fragments (**a**, **b**) indicating that during the process, isolated long RNAs underwent re-cleavage. For sequence analysis, bands **c***, **d*** and **e*** were excised, the DNA was eluted, pooled and cloned as described in section 1.5. Note that the designation of bands **a** to **e** used here does not correspond to bands **a** to **e** discussed in the main text.

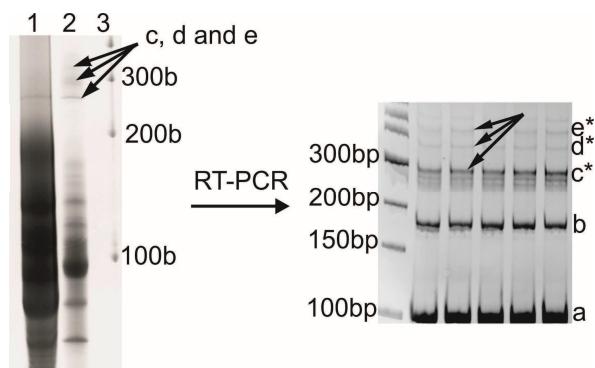


Figure S1: Left: *In vitro* transcription and self-processing of CRZ-2 analyzed in a 15% denaturing polyacrylamide gel corresponding to Fig. 3 in the main text. Lane 1: *in vitro* transcription mixture, lane 2: incubation of the isolated full length transcript (103mer) in ribozyme reaction buffer (10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 37 °C, 2 h (cleavage) and 50 mM MgCl₂ 37 °C 2 h (ligation)), lane 3: RNA ladder. Right: 10% native polyacrylamide gel of the amplified cDNA mixture obtained after RT-PCR of RNA fragments **c**, **d** and **e** resulting from self-processing of the linear 83mer (lane 1 on the left). Lane 1: DNA size marker, lane 2-6 amplified cDNAs: **a**: DNA product, including primer sequences, corresponding to 83mer RNA, **b**: DNA product, including primer sequences, corresponding to RNA dimers, **c**, **d**, and **e**: isolated RNAs, **c***, **d*** and **e*** corresponding DNAs.

2. Concentration dependence of concatemerization

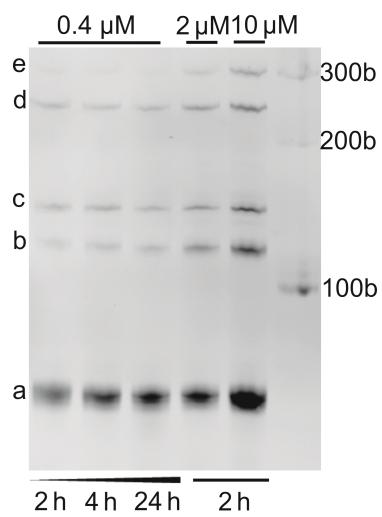


Figure S2: 15% polyacrylamide gel analysis of re-incubation of the isolated linear 83mer in ligation buffer for 2 to 24 h at different concentrations. Right lane: RNA ladder. Product band intensities were quantified using Gene ImageIR 4.05, and the relative amount of each species within one lane was calculated.

Sequence-controlled RNA self-processing: computational design, biochemical analysis and visualization by AFM

Sonja Petkovic^{1,#}, Stefan Badelt^{2,#}, Stephan Block³, Christoph Flamm², Mihaela Delcea³, Ivo Hofacker^{2,4}, Sabine Müller¹

¹Institute for Biochemistry, Ernst-Moritz-Arndt University Greifswald, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany

²Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17/3, A-1090 Vienna, Austria

³ZIK HIKE - Center for Innovation Competence, Humoral Immune Reactions in Cardiovascular Diseases, Ernst-Moritz-Arndt University Greifswald, Fleischmannstraße 42-44, -17489 Greifswald, Germany

⁴Research Group Bioinformatics and Computational Biology, University of Vienna, Währingerstraße 29, A-1090 Vienna, Austria

shared first authorship

Corresponding authors:

Prof. Dr. Sabine Müller

Institute for Biochemistry, Ernst-Moritz-Arndt University Greifswald, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany,

Email: smueller@uni-greifswald.de

Prof. Dr. Ivo Hofacker

Institute for Theoretical Chemistry and Structural Biology, University of Vienna, Währingerstraße 17/3, A-1090 Vienna, Austria,

Email: ivo@tbi.univie.ac.at

Dr. Stephan Block

ZIK HIKE, Ernst-Moritz-Arndt University Greifswald, Fleischmannstraße 42-44, -17489 Greifswald, Germany; current address: Department of Applied Physics, Chalmers University of Technology, Fysikgränd 3, S-412 96 Gothenburg, Sweden,

Email: stephan.block@chalmers.se

ABSTRACT

Reversible chemistry allowing for assembly and disassembly of molecular entities is considered important for biological self-organization. With regard to the RNA world theory, ribozymes supporting both cleavage and ligation of phosphodiester bonds may have contributed to enhancing the complexity of RNA molecules and thus, to extend the functional space. In modern cellular life, RNA processing plays a vital role, and even self-processing activities are still found in RNAs occurring in a variety of viruses and viroids. We have engineered by computer-aided design a number of self-processing RNAs derived from the hairpin ribozyme. Self-processing starts from RNAs produced by *in vitro* transcription and incubated at conditions typical for hairpin ribozyme cleavage/ligation assays. Reactions involve two cleavage events to remove 5'- and/or 3'-terminal sequence patches, followed by spontaneous intra- and intermolecular ligation of the produced fragments. Circular RNAs and concatemers are the typical products. The cyclic or concatemeric nature of RNA species was confirmed by gel electrophoresis, and individual species were visualized at the single-molecule level by AFM. The results of this study underscore the likelihood of RNA as a major player in early life forms, and moreover, demonstrate the power of an interdisciplinary effort in solving complex design problems.

INTRODUCTION

RNA processing plays a fundamental role in the cellular life cycle. RNA molecules are permanently synthesized, modified, edited, truncated or abolished. In viruses, viroids and satellite RNAs with circular RNA genomes, replication follows a rolling circle mechanism, thus initially producing linear concatemeric versions of the RNA genome(1). Further processing is required to convert the concatemers back to monomers that subsequently are cyclized to yield the final replication product: a cyclic RNA complementary to the template. This processing is dependent on specific RNA structural motifs that support reaction at the site of cleavage and ligation (2,3). Among those, the hammerhead and the hairpin ribozyme are probably the best studied small RNAs with catalytic activity(4,5). We are interested in the hairpin ribozyme and its structural manipulation for functional control, and we have engineered a number of hairpin ribozyme variants that allow for tuning of cleavage and ligation activity (6-13). A recent example is a self-processing RNA molecule that, analogous to its viral origin, runs through a cascade of cleavage and ligation reactions(6,14). Activity of

this engineered self-processing RNA, termed CRZ-2, was analyzed by one- and two-dimensional polyacrylamide gel electrophoresis, exoribonucleolytic decay and Sanger sequencing(6). As a result, we have confirmed that cyclic and linear products (concatemers) are formed. Self-processing activity is derived from the natural satellite RNA activity that is cleavage and ligation via a transesterification mechanism(15), and is applied for engineering of hairpin ribozyme variants such as CRZ-2. Cleavage occurs by nucleophilic attack of the 2'-oxygen on the neighboring phosphorous resulting in a trigonal-bipyramidal intermediate. Upon release of the 5'-OH-group, a 2',3'-cyclic phosphate is formed. Ligation follows the same reaction path in opposite direction and proceeds *via* ring opening of the cyclic phosphate, exclusively delivering the natural 3',5'-phosphodiester (Fig. 1).

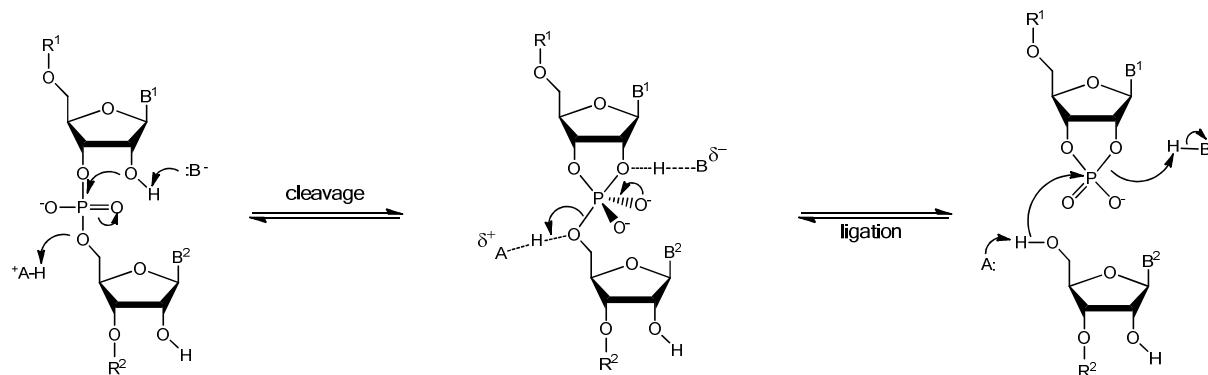


Figure 1: Hairpin ribozyme mechanism (from left to right) nucleophilic attack followed by intermediate formation and release of newly formed 5'- and 3'- termini (see text for details); R^1, R^2 = oligonucleotide chain, B^1, B^2 = nucleobase, A = acid, B^- = base.

Ligation is favored when fragments or termini are tightly bound to the ribozyme. In this case the favorable enthalpy of ligation compensates for the anyway low entropic cost. On the contrary, if cleavage products/ligation fragments are less tightly bound, dissociation becomes a major factor. In this case, the entropic cost of ligation is strongly increased, and cleavage becomes the favorable reaction(16). These characteristic features distinguish the hairpin ribozyme from other endonucleolytic ribozymes and allow for control of the cleavage and ligation activity(6,8,12,14). The previously designed self-processing hairpin ribozyme variant CRZ-2 is a bi-stable RNA. It was engineered to fold in two alternative cleavage favoring conformations (Fig. 2a), from which first the 5'- and then the 3'-terminus or *vice versa* is cleaved off. The remaining shortened fragment favors ligation, because the 5'- and 3'-end with hydroxyl group and cyclic phosphate are closely pre-oriented and cannot freely dissociate. Intramolecular ligation delivers a cyclic RNA species. In addition, we have shown that as a

result of self-association and intermolecular ligation also linear concatemers are formed(6) (Fig. 2b).

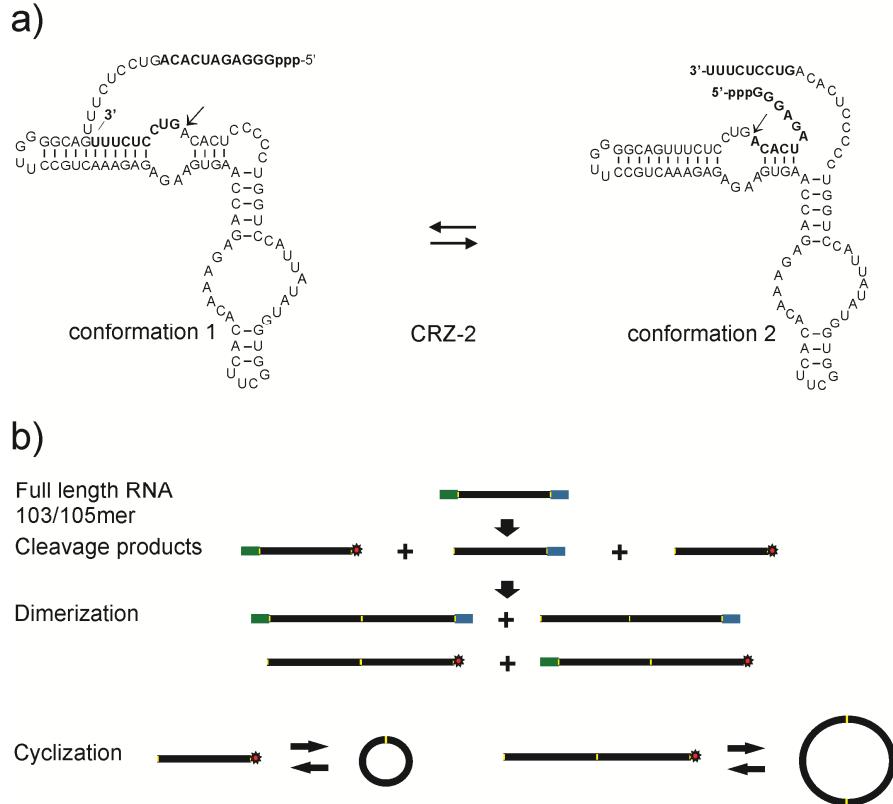


Figure 2: **a:** Two alternative cleavage-favoring conformations of the reference self-processing RNA CRZ-2, **b:** schematic view of self-processing reactions.

Here we used computer-aided design to synthesize new self-processing hairpin ribozyme variants in order to better control activity and to favor either intermolecular or intramolecular reactions. Furthermore, we complemented our biochemical analysis of self-processing product ensembles by single molecule visualization techniques. We present high resolution AFM images of the reaction mixtures, visualizing even the rather short 83mer RNA fragment.

EXPERIMENTAL SECTION

Computational Ribozyme Design

In order to have a consistent, length-independent annotation for all possible RNA species that can emerge from a starting (full-length) ribozyme, we introduce the following notation: We denote the 5'- and 3'-ends of the full-length molecule as *L* (left) and *R* (right), respectively, and the linear core as *C*(center). An initial ribozyme species therefore consists of three parts and can be abbreviated as “LCR” molecule. Additionally we introduce the term *O* for the

circular version of C . Despite the ability of C to form a circular O , multiple copies of C can ligate to one long strand that itself can form a maxi-cycle (e.g. $CCC \leftrightarrow C_3 \leftrightarrow O_3$).

The following two scoring functions (κ_1 and κ_2) were used to select for ribozymes which are able to process themselves efficiently into cyclic monomers (κ_1) and to differentiate between those, which predominantly form catalytically active or inactive dimers (κ_2). Both functions estimate rates for cleavage reactions by computing the probabilities of catalytic secondary structures, hence following two hypotheses: first, a cleavage/ligation rate is proportional to the equilibrium probability of a catalytically active secondary structure; second, the cleavage reaction leads to dissociation of the shorter fragment and is therefore irreversible. Equilibrium probabilities of RNA molecules can be calculated from the equilibrium partition function (Z); Z can be calculated using the McCascill algorithm (17) implemented in RNAfold of the Vienna RNA package (18). Let $Z(LCR)$ be the equilibrium partition function over all feasible secondary structures compatible with the molecule LCR , and $Z(LCR^L)$ be the constraint partition function over all reactive secondary structures in which L can be cleaved off, then the probability $P(LCR^L)$ can be computed as

$$P(LCR^L) = \frac{Z(LCR^L)}{Z(LCR)} \quad [1]$$

All computations were done using the Vienna RNA package Version 2.1.6 with standard energy parameters at 37°C. Figure S1 shows our model of the cleavage cascade yielding cyclic monomers. It starts with a full length molecule (LCR) that can process itself into the linear catalytic core in two parallel ways. Either the 5'-end (L) of the sequence is cleaved first and the resulting truncated version (CR) cleaves the 3'-end (R), or *vice versa*. For both of these parallel, two-step reaction pathways we are interested in the rate limiting step which determines the speed of the cascade. Since we approximate cleavage rates from probabilities of catalytic secondary structures, the rate limiting cleavage reaction is the minimum of both probabilities, and the total rate is the sum of both parallel cleavage pathways. The yield of circular reaction products is computed as the fraction of circular molecules in equilibrium $\left(\frac{Z(O)}{Z(O)+Z(C)}\right)$ resulting in the following scoring function:

$$\kappa_1 = \left(\min \left\{ \frac{P(LCR^L)}{P(CR^R)}, \frac{P(LCR^R)}{P(CR^L)} \right\} + \min \left\{ \frac{P(LCR^L)}{P(LC^L)}, \frac{P(LCR^R)}{P(LC^R)} \right\} \right) \times \frac{Z(O)}{Z(O)+Z(C)} \quad [2]$$

Our model of the cleavage/ligation cascade which forms circular dimers is shown in Figure S1b. It follows the assumption that dimerization between full-length molecules happens first, then an intramolecular cleavage cascade is triggered, and finally the system equilibrates between all dimeric cleavage products. While monomers have one reactive ground state with two conserved interior loops to cleave one of their ends, dimers can form up to two reactive centers in three different ways to cleave one end (see Figure S1b). The two interior loops needed for a reaction are commonly called loop A (harboring the reactive site) and loop B (5). Our computations to score the dimer–cleavage cascade require at least one of these loop regions to be formed intermolecularly, since κ_1 already scores the molecules according to their intramolecular cleavage efficiency. The probability to cleave both 5'-ends (L) from a LCR dimer $P(LCR_d^{2L})$ can therefore be computed as

$$P(LCR_d^{2L}) = \frac{Z(LCR_d^{2L})}{Z(LCR_d)} \quad [3]$$

where $Z(LCR_d^{2L})$ is the sum of two distinct sets of structures in which loop B is either formed intramolecularly or intermolecularly.

Similar to κ_1 , the yield of circular dimers is computed as the fraction of circular dimers in equilibrium $\left(\frac{Z(O_2)}{Z(CC) + Z(C_2) + Z(O_2)} \right)$ with CC and C_2 denoting non-covalently and covalently bound dimers, respectively. The second scoring function κ_2 is therefore computed as

$$\kappa_2 = \frac{[LCR_d]_\theta}{[LCR]_\theta} \left(\min \left\{ \frac{P(LCR_d^{2L})}{P(CR_d^{2R})}, \frac{P(LCR_d^{2R})}{P(LC_d^{2L})} \right\} \right) \frac{Z(O_2)}{Z(CC) + Z(C_2) + Z(O_2)} \quad [4]$$

where the first term $\frac{[LCR_d]_\theta}{[LCR]_\theta}$ computes the equilibrium ratio between dimers and monomers at a given concentration θ (here 100nM) for the LCR molecule following Bernhart *et al.*(19). The scoring function only maximizes the probabilities for catalytically active homo-dimers; pathways that involve dehybridization of partially cleaved species are not included. This corresponds to the assumption that intramolecular cleavage reactions as well as intramolecular folding kinetics are faster than intermolecular structural rearrangements.

Design of an inactive cyclic dimer

Circular monomers have been successfully identified for CRZ-2 in our previous work(6); however, the formation of circular dimers could not be shown. Since the size of cyclic RNAs is not assessable in PAA gels according to standard size markers, we designed an inactive

circular dimer (CRZ^*) as a reference. The cyclic reference dimer should be as similar as possible to the circular dimer of CRZ-2 and therefore help to identify this species in a PAA gel. However, due to its inactivity, the linear version of the reference dimer (produced by *in vitro* transcription) has to be ligated enzymatically to the desired cyclic product.

In particular, the design of an inactive circular dimer has to fulfill the following constraints with respect to the reference circular dimer (CRZ): (i) the secondary structure ensemble associated with the RNA has to be similar, (ii) the nucleotide content has to be equal, (iii) the sequence must not be symmetric, (iv) all conserved catalytic centers have to be destroyed and (v) a T7 promotor region is needed. Points (i) and (ii) shall insure a similar migration pattern on a polyacrylamide gel, whereas points (iii), (iv) and (v) are necessary for experimental implementation. Asymmetry insures that only the defined 3'-terminal regions of Klenow primers overlap to obtain specific dsDNA of the desired length as template for RNA synthesis, and inactivity is necessary to avoid cleavage/ligation reactions after and during *in vitro* transcription. As a first step we preset the residual T7 RNA promotor sequence 5'-GGG AGA-3' as a non-mutable hexanucleotide at the 5'-end of the ribozyme. These bases will inevitably occur in the *in vitro* transcribed RNA due to usage of T7 RNA polymerase. Since this pre-processing step harms condition no. (ii), we mutated different helical regions to compensate for inequalities in the nucleotide content. Next, we randomly flipped base pairs within all helical regions (apart from the residual T7 RNA promotor region) and randomly shuffled the nucleotides from all loop regions to obtain loss of catalytic activity. With this approach we designed roughly 500 RNA species that fulfill conditions (ii), (iii), (iv) and (v). To ensure similar folding behavior, the sequence should not only have the same ground state as the CRZ-2 dimer, also the whole structure ensemble should be similar. Therefore, we first selected for those sequences that have the smallest mean base pair distance within the equilibrium structure ensemble. This mean base pair distance (D) can be computed as

$$D(CRZ, CRZ^*) = \sum_{ij} P_{ij}^{CRZ} (1 - P_{ij}^{CRZ^*}) + P_{ij}^{CRZ^*} (1 - P_{ij}^{CRZ}) \quad [5]$$

With P_{ij}^{CRZ} denoting the probability of a single base pair between position i and j for the molecule CRZ . From the top 20 designed molecules, we selected a sequence (shown in Table S1, Supporting Information) that has a comparable minimal free energy (MFE).

RNA preparation and analysis - General remarks and chemicals

Deoxynucleotidetriphosphates (dNTPs), nucleotide triphosphates (NTPs), Klenow buffer, DNase I, T7 RNA polymerase, Klenow fragment exo-, RiboLock™, RiboRuler™ low range RNA ladder and polynucleotide kinase were purchased from *Fermentas Company* (Schwerte, Germany); T4 RNA Ligase 2 (T4 RnL2) and the appropriate buffer was obtained from *New England Biolabs*(Frankfurt am Main, Germany). DNA primers were provided by *Biomers.net* (Ulm, Germany).RNase R including the required buffer was obtained from *epicenter* (Oldendorf, Germany). All chemicals and reagents were of analytical grade and filtered through a 0.2 µm polyvinyl difluoride membrane before use. Upon electrophoresis, all gels were stained for 5 to 10 min with ethidium bromide. Final concentration of ethidium bromide in 1xTBE was 0.5 µg/ml.All UV spectra were recorded on a NanoDrop ND 1000 spectrophotometer. Stained gels (agarose or polyacrylamide) were visualized using Chemi-Smart 2000 WL/LC 26M or VWR GenoView.

RNA generation

Klenow primers (for sequences see Table S1 in Supporting Information) with 20 bp overlap (27 bp for the inactive dimer) were used in Klenow reactions with Klenow exo⁻polymerase following the manufacturer's protocol, and stopped by precipitation from ethanol at -20 °C overnight. DNA was isolated from native agarose gels (1.5%, EtBr stained). Product containing bands were cut out and DNA was isolated using QIA quick gel extraction kit (*Qiagen*, Venlo, The Netherlands). Since only one product was detectable after Klenow reaction, in later preparations the gel extraction step was skipped. Instead, after ethanol precipitation of the Klenow reaction product, the pellet was solved in 100 µl water, and 5 µl were used for subsequent *in vitro* transcription.

RNAs were synthesized by *in vitro* transcription of double stranded DNA templates (1 µM concentration, or as mentioned above 5 µl of the Klenow product resolved in water after precipitation) with T7 RNA polymerase in the presence of the four ribonucleoside triphosphates (2 mM) and 1 U/µl RiboLock™ in 1x HEPES buffer (Na-HEPES 50 mM, MgCl₂*6H₂O 12 mM, Spermidin 2 mM, pH = 7.5) in a total reaction volume of 50 µl for 3 h at 37 °C. DNA template was hydrolyzed adding 2 µl DNase I directly to the transcription mixture and left at 37 °C for additional 30-45 min. Final purification was achieved by electrophoresis on 15% denaturing polyacrylamide gels (for composition see subchapter *PAGE analysis* below), elution of the product-containing bands with sodium acetate (0.3 M, pH= 7, 3 times for at least two hours and overnight for the final elution step, shaking at

approximately 500 rpm, at 10 °C) and precipitation with 250 vol.-% ethanol at -20 °C overnight.

Preparation of the inactive cyclic dimer by transcription priming with GMP

To obtain the inactive linear dimer (*in-l-166mer*), with 5'-terminal monophosphate, GMP was added to the NTP mix following the protocol of Harris and Christian for incorporation of guanosine monophosphorothioate(20). A 4.8:1 ratio of GMP:GTP was used, and the double stranded Klenow DNA, buffer, RiboLock™ and polymerase were added as described above. *In vitro* transcription was stopped after 3 hours at 37 °C, and double stranded DNA template was hydrolyzed using DNase I following manufacturer's protocol. The reaction mixture was blended with 100 vol-% stop mix (7 M urea and 50 mM EDTA) and directly used for purification on a 15% denaturing polyacrylamide gel. After PAGE, elution of the desired RNA and ethanol precipitation as described above, RNAs were used for ligation.

Enzymatic ligation in the double stranded region of the *in-l-166mer* to generate the cyclic species *in-c-166mer* was conducted using T4 RnL2 in a total reaction volume of 20 µl at 37 °C for 4 hrs following the suppliers protocol. RNA was purified using the RNA Clean & Concentrator™-5 kit (ZymoResearch, Freiburg, Germany) following the general protocol for total RNA purification. Elution of RNA was carried out with 50 µl desalting and purified millipore water. After addition of 50 µl stopmix, ligation products were analyzed on a 15% denaturing polyacrylamide gel.

Exoribonucleolytical decay

10 µl of T4 RNA ligase 2 reaction mixture were used directly for hydrolysis using RNase R. MgCl₂ to a final concentration of 5 mM, RNase R buffer and water up to 17 µl were mixed and denatured at 90 °C for 5 min. The mixture was cooled down to 50 °C for 1 min before addition of 1 µl of an 1:1 freshly with water diluted RNase R solution. Hydrolysis occurred at 50 °C for 10 min. Reaction was stopped using an equal volume of stop mix, which is also used as loading buffer for electrophoresis, and the mixture was immediately frozen in liquid nitrogen. Reaction products were analyzed by electrophoresis through a 15% denaturing polyacrylamide gel.

Self-processing reactions

RNAs (11.25 pmol) were solved in Tris-HCl buffer (10 mM, pH= 7.5). After denaturation for one minute at 90 °C, RNA folding was allowed for 10 min at room temperature. To initiate

the cleavage reaction, MgCl₂ hexahydrate to a final concentration of 10 mM was added and reaction was allowed to proceed for 2 h at 37 °C. To favor ligation, Mg²⁺ concentration was increased up to 50 mM, and reaction proceeded for additional 2 h at 37 °C. Reaction was stopped using an equal volume of stop mix composed of urea (7 M) and EDTA (50 mM) for the following PAGE analysis.

PAGE analysis

For RNA species analysis or purification, denaturing (7 M urea) polyacrylamide gel electrophoresis (acrylamide: bisacrylamide 19:1 100 ml, ammonium persulfate 10 % w/v, 1 ml, *N,N,N',N'*-Tetramethylethane-1,2-diamine 50 µl) was applied, using 1xTBE buffer as running buffer and stop mix (7 M urea, 50 mM EDTA, bromophenol blue and xylene cyanol each 5 vol-%) for sample loading. After mixing samples and/or RNA size standard with buffer, RNAs were denatured at 90 °C for 2 min and directly loaded onto the gel. Loading buffer for the RNA size standard was provided by *Fermentas* (Schwerte, Germany).

Two-dimensional electrophoresis

Identification of circular RNAs by 2D electrophoresis is based on the fact that the migration of linear and circular nucleic acids is distinctly dependent on the gel pore size(21,22). To enrich the samples with linear RNAs for better identification of the circular species, a commercially available RNA size standard (RiboRuler™ low range RNA ladder; *Fermentas*) being composed exclusively of linear RNAs was added. Each individual mixture was separated in the first dimension gel. Then, the gel piece corresponding to the entire lane was cut out and used for electrophoresis in the second dimension, upon which linear RNAs are supposed to form a diagonal. Covalently closed cyclic RNAs possess reduced degrees of freedom, thus migrating in non-linear dependence on the linear species and occurring beyond the diagonal (23).

All ribozyme variants (11.25 pmol) were analyzed using two-dimensional PAGE (for polyacrylamide gel composition, buffers and staining see above *PAGE analysis*. First dimension: denaturing conditions (7 M urea) 15% polyacrylamide; second dimension: 17.5% denaturing polyacrylamide or 15 % native polyacrylamide.

RNA preparation for AFM analysis

Ribozyme reactions were carried out as described above using 400 nM RNA. After reaction, the product mixture was diluted 1:10, and 5 µl of this solution were lyophilized. The pellet

was taken up in 50 µl of semi-denaturing buffer (25 mM EDTA, 3.5 M urea) to a final RNA concentration of 4 nM for imaging. Re-solved RNA samples were frozen in liquid nitrogen until use.

Atomic force microscopy (AFM)

AFM imaging was performed in air using a Multimode atomic force microscope (*Veeco/Digital Instruments*, Santa Barbara, CA) equipped with a NanoscopeIIIa controller. The AFM piezo scanner was calibrated using calibration gratings TGZ01 (rectangular 26 nm SiO₂ steps on silicon wafer; *MicroMasch*, Tallin, Estonia) and PG (chessboard like pattern on silicon, 100 nm depth and 1 µm pitch; manufacturer: *Digital Instruments*, Santa Barbara, CA). RNA samples were prepared by placing a small droplet of RNA solution onto freshly cleaved mica (*SPI Supplies*, West Chester, PA). For the investigated RNA constructs, adsorption times of 30 seconds to 2 minutes were sufficient to obtain a suitable RNA surface coverage on the mica substrate. After adsorption, the RNA samples were rinsed in Milli-Q water (*Millipore*, Billerica, MA) and dried in a laminar flow hood, followed by AFM imaging. The images were recorded with conventional Tapping Mode in air using standard tapping mode cantilevers (OMCL-AC160TS, *Olympus*, Hamburg, Germany). Before usage the cantilevers were tested with a Nioprobe self-imaging sample (*Aurora Nanodevices*, Nanaimo Canada) and only cantilevers with tip radius < 5 nm were used for imaging. To reduce tip contamination by RNA uptake during imaging process, cantilevers were functionalized with 3-aminopropyltrimethyl-ethoxysilane (APDES) from *ABCR* (Karlsruhe, Germany) one day prior usage.

In contrast to DNA samples, whose structure often remains unchanged even after storage periods of several months (as judged by their spatial properties in AFM imaging), samples had to be imaged within few days after preparation. The highest resolutions were always obtained directly after preparation, while storing in air often led to post-preparational RNA chain aggregation already after few weeks.

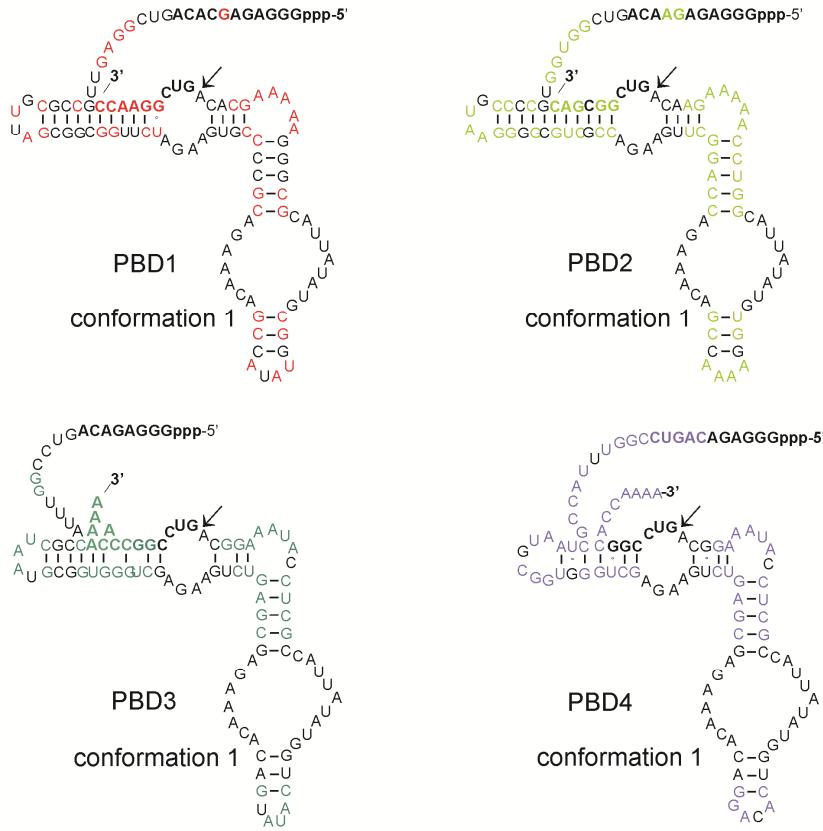
Images were analyzed using the software supplemented with the AFM. The shape of an RNA chain was ‘retraced’ in terms of a sequence of connected straight segments, which allowed to calculate the contour length as sum of Euclidean distances(24). As shown by Rivetti and co-workers (who numerically assessed the accuracy of different methods for contour length determination) this approach has an intrinsic error < 1%(25). The main source of error in the contour length determination is therefore given by the lateral resolution of the AFM, which is on the order of few nanometers (see Discussion section for details).

RESULTS

Computer-aided sequence design

Compared with manual design that we had applied in previous work (6,14), computer-aided design is a more sophisticated way towards control of self-processing activity of RNA species. Therefore, we have started a bioinformatics approach to evolve hairpin ribozyme derived RNAs with self-processing activity. We have designed two classes of ribozyme species: members of the first class should process themselves efficiently into circular monomers, whereas members of the second class would maximize the yield of ligation competent dimers, hence forming linear and circular dimers that can further concatemerize. The design process is complicated by the fact that multiple constraints exist on both sequence and structure level. On the sequence level we included two well-conserved interior loop regions from the hairpin ribozyme (26), as well as a T7 RNA promotor sequence at the 5'-end for experimental implementation. On the structural level, the constructs have to form two distinct catalytic centers to cleave off both the 5'- and 3'-ends as depicted in Figure 3. Our approach is a two-step process that first computes a large set of RNA sequences with catalytic properties, and second scores these sequences to select for ribozymes with the desired behavior. Previously, we have shown that the efficient design of bi-stable molecules is surprisingly easy (27). The algorithm, implemented in the program switch.pl of the Vienna RNA package, mutates initially random sequences into bi-stable switches *via* consistent mutations guided by a dependency graph. The mutations are meant to increase the probability of forming catalytically active structures and influence the conformations formed upon dimerization of the individual species. Using switch.pl we designed roughly 10000 bi-stable RNA molecules conforming to the above design objective. We used two scoring functions (κ_1 , κ_2) to rank molecules according to their probabilities of forming reactive structures. We applied this approach to our reference system CRZ-2 and related the scoring functions of test systems. A less negative κ_1 -value indicates high catalytic activity of all monomeric variants and therefore is maximized for all selected ribozymes. In contrast, a less negative κ_2 indicates a high probability to form catalytically active homo-dimers, hence it was used to discriminate between ribozymes that are meant to favor formation of cyclic dimers (less negative κ_2) and those that do not (more negative κ_2). Detailed explanation and formulas can be found in the Experimental Section and in the Supporting Information (Figures S1a and S1b). Figure 3 shows four bi-stable ribozyme sequences (PBD1 to PBD4) that were selected for experimental validation in comparison to the reference system CRZ-2.

a)



b)

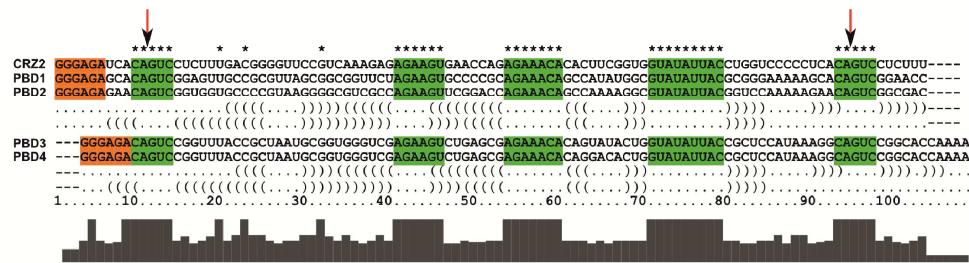


Figure 3: **a:** Secondary structures of PBD 1 to 4 in conformation 1. PBD1 to 4 are bi-stable and may fold in an alternative second conformation as shown for CRZ-2 in Figure 2a. Sequence changes in comparison to the reference RNA CRZ-2 are colored. **b:** Sequence alignment of the four designed RNAs PBD1 to 4 with the reference system CRZ-2. Green interior loop areas are reported to be essential for cleavage/ligation activity and were therefore fixed during the design process. The orange colored T7 RNA promoter sequence was needed for experimental implementation. The secondary structure in dot-bracket notation below shows the constraints on a structural level.

Table 1 summarizes their expected properties and the results from the scoring functions (rounded to two decimal figures). PBD1 and PBD2 have the same length of all cleavage products compared to CRZ-2. PBD3 and PBD4 have differently sized 5'- and 3'-ends. The last two columns show the logarithm of the results from our scoring functions (see equations 2 and 3 in the Experimental Section).

Table 1: Summarized properties of the designed sequences.

RNA	Full length	Fragment length	Log(κ_1)	Log(κ_2)
CRZ-2	103	11+83+9	-19.24	-37.58
PBD1	103	11+83+9	-9.38	-13.71
PBD2	103	11+83+9	-9.84	-28.71
PBD3	105	8+83+14	-12.14	-18.77
PBD4	105	8+83+14	-12.13	-31.23

Compared to reference RNA CRZ-2, PBD1 to 4 differ by base replacements in all non-conserved regions (Fig. 3). However, the new designed ribozymes were meant to undergo the same cleavage cascade reaction as described for CRZ-2 previously(6) and depicted in Figure 2b. Dimerization of the hairpin ribozyme was demonstrated previously(28), and we have discussed this feature as an essential prerequisite for the formation of concatemers by CRZ-2. Therefore, sequences forming catalytically active, intermolecular ligation competent dimers, in our design approach are assumed to favor concatemerization (PBD1 and PBD3), whereas sequences that have lower tendency to form these structures, are assumed to predominantly form cyclic monomers (PBD2 and PBD4). Our scoring functions indicate that PBD1 and PBD2 show increased efficiency to form cyclic monomers (Table1, log (κ_1)) compared with PBD3 and PBD4. PBD1 and PBD2 were designed with the same constraints as CRZ-2, which enabled optimization of these ribozymes to fold likewise in both catalytically active conformations; therefore extending the pool of designed species with highly efficient ribozymes that would show exactly the fragments as reference system CRZ-2.

Analysis of the self-processing behavior of the designed sequences

The five RNAs, CRZ-2 and PBD1 to 4 were prepared by *in vitro* transcription with T7 RNA polymerase and incubated at conditions favoring self-cleavage followed by ligation (see Experimental Section and (6)). Reaction products were analyzed using denaturing

polyacrylamide gels (Figures 4,5,6) and visualized by AFM (Figures 7 and 8). Table 2 shows the lengths (in number of bases) of products that theoretically can be formed upon reaction.

Table 2. Lengths of possible cleavage and ligation products upon RNA self-processing

RNA	Full length (a) ¹⁾	Cleavage products (b, c) ¹⁾	Ligation products composed of only 83mers (d+e) ^{1), 2)}	Ligation products of mixed composition (e) ^{1), 2)}
			(d+e) ^{1), 2)}	(e) ^{1), 2)}
PBD1	103	83 (c), 92, 94 (b)	c83 (d), 166, c166, 249 (e)	175, 177, 186
PBD2	103	83 (c), 92, 94 (b)	c83 (d), 166, c166, 249 (e)	175, 177, 186
PBD3	105	83 (c), 91, 97 (b)	c83 (d), 166, c166, 249 (e)	174, 180, 188
PBD4	105	83 (c), 91, 97 (b)	c83 (d), 166, c166, 249 (e)	174, 180, 188
CRZ-2	103	83 (c), 92, 94 (b)	c83 (d), 166, c166, 249 (e)	175, 177, 186

¹⁾a: full length transcript, b: cleavage intermediates, c: final cleavage product, d: cyclic monomer, e:concatemers, compare also legend to Figure 4

²⁾ Note that in addition to dimers and trimers also longer concatemers can be formed.

Figure 4 shows an overview of reactions of all self-processing ribozymes (CRZ-2, PBD1-4). For comparison, the linear 83mer (l-83mer) resulting from two cleavage events in CRZ-2 and being incapable of further cleavage was incubated at identical conditions (Fig. 4, lane 7). The behavior of this 83mer was analyzed in detail recently(6); such that the band pattern produced by the l-83mer could be used as guideline to navigate through the PAA gel and to assign the obtained bands to individual RNA species. This becomes especially important, since chemical modifications at RNA ends (such as OH, phosphate or cyclic phosphate), RNA sequence itself, and RNA structures formed in spite of denaturation can affect the migration behavior of RNA molecules. Thus, the standard length marker (lane 2) can only serve as an approximate guideline for higher ligation products.

Full length transcripts (a – 103/105mers)

The 103- (CRZ-2, PBD1 and PBD2) or 105mers (PBD3 and PBD4) are typically located below the 100 nt size standard mainly due to the triphosphate at the 5'-end resulting from *in vitro* transcription of the ribozymes (Figure S3). The 103mer of CRZ-2 and PBD1 is barely detectable after ribozyme reaction (lane 1 and 3), whereas full length transcripts of PBD2 to PBD4 (lanes 4, 5 and 6) are still visible. This implies that CRZ-2 and PBD1 exhibit higher activity in cleaving off the 5'- or the 3'-endor both, and producing the shortened fragments denoted with **b** and **c** (Table 2, Fig. 4).

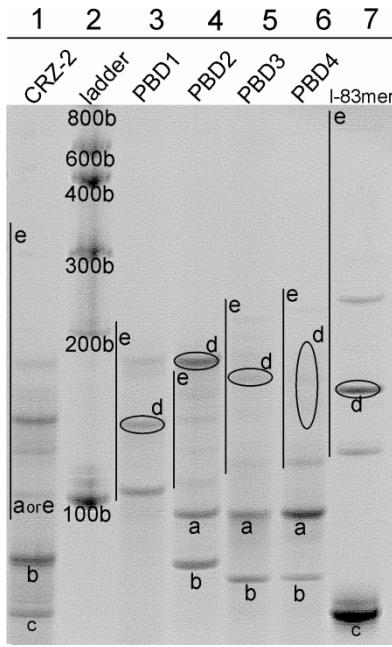


Figure 4: Analysis of self-processing reactions of sequences PBD1-4, CRZ-2 and the linear 83mer (l-83mer) in a 15% denaturing polyacrylamide gel, lane 2: RNA size standard. Bands were assigned as follows: full length transcripts (**a**); cleavage intermediates (5'- or 3'-truncated transcripts) (**b**); final cleavage product (5'- and 3'-truncated transcripts) (**c**); cyclic RNA resulting from intramolecular ligation of species **c**(**d**); concatemers resulting from intermolecular ligation of species **b** and **c**(**e**). See also Figure S2 (Supporting Information).

Cleavage products (**b** – 97, 94, 92, 91mer and **c**- linear 83mer)

In CRZ-2, PBD1 and PBD2, a 92mer and a 94mer are produced as intermediates upon the first cleavage. These two intermediates occur as one band, since the 94mer carries additional charges from the triphosphate at the 5'-end. For CRZ-2 and PBD2 (Fig. 4, lanes 1, 4), a prominent 92/94mer band is visible, PBD1 shows none of these species. PBD3 and PBD4 produce a 91mer and a 97mer, with the 91mer carrying the triphosphate. Both systems show 91mers, whereas 97mers are not or barely detectable (lanes 5, 6). However, it should be noted that at higher concentrations, also the 97mer was detectable for PBD4 (see Supporting Information Fig. S2, panel d).

The final cleavage product of all test systems is a linear 83mer. Figure 4 (lane 7) shows the 83mer from CRZ-2 used as an additional size standard (see (6) for behavior of the l-83mer from CRZ-2). Interestingly, only CRZ-2 shows a lane of the final cleavage product, while

83mers of PBD1-4 are not detectable, suggesting an immediate consumption in ligation reactions.

*Intramolecular monomeric ligation (**d** – cyclic 83mer)*

From all produced monomers (l-83mer, 91mer, 92mer, 94mer, 97mer, 103mer and 105mers) the linear 83mer is the only RNA that may perform cyclization due to its chemical constitution at its 3'- and 5'-end. However, the migration behavior of an unknown cyclic species in a PAA gel is impossible to predict by common size markers. The overall shape and the migration behavior of the cyclic RNA strongly depend on the sequence (21,29). Analysis by ligation site detection with labeled oligonucleotide probes and thus identification of ligation products is not possible, because such an approach would label every possible cyclic and linear ligated species except of the linear 83mer. Isolation of a single RNA species, i.e. the assumed cyclic monomer, from the gel for analysis and identification also falls apart, since the isolated species would undergo self-processing during isolation and purification, and hence produce again a multitude of RNA species (a fact that we made use of regarding the l-83mer marker).

Previously, we have set up a 2-dimensional-PAA gel electrophoresis assay (see Experimental Section and (6)) to identify cyclic species by means of their non-linear movement at different PAA concentrations. For CRZ-2, we have shown that the cyclic 83mer is located approximately at the 150 nucleotide size standard; for PBD1-4, identification of cyclic species by 2D gel electrophoresis can be seen in Figure 5. While linear species move on a diagonal in the second dimension, cyclic species are expected to show irregular movement. According to the results shown in Figure 4, full-length CRZ-2 (lane 1) does not form a circular 83mer, while incubation of the isolated linear 83mer of CRZ-2 alone (lane 7) clearly produces the cyclic species. Interestingly, all newly designed RNAs (PBD1-4, lanes 3-6) do show circular 83mers, while not showing any linear 83mers. In case of PBD4, the cyclic species is not represented by a discrete band, but rather appears as a smear.

*Higher non-cyclic ligation products (**e** – dimers, trimers, concatemers)*

Intermolecular backbone ligation can only occur upon dimerization of the 83mer and/or the intermediate cleavage products, carrying the required termini (5'-OH and 2',3'-cyclic phosphate). A summary of these species can be seen in Table 2.

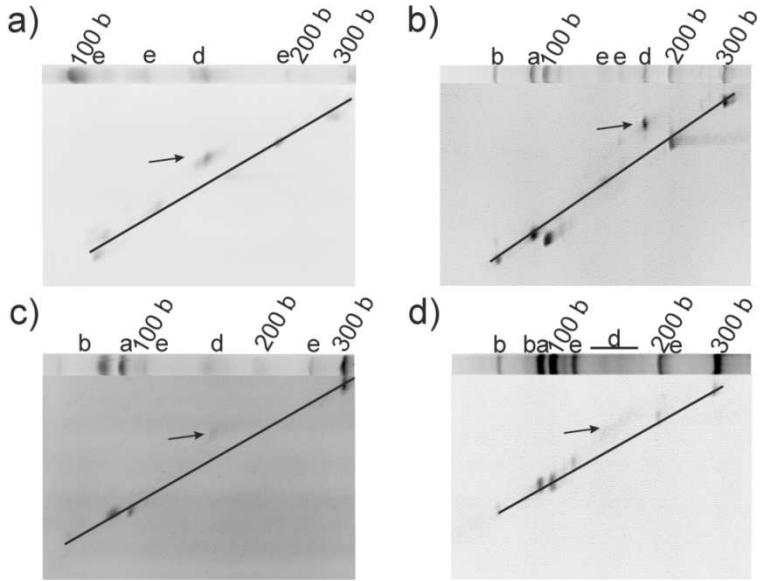


Figure5: 2D gel electrophoretic analysis of PBD1 (**a**), PBD2 (**b**), PBD3 (**c**) and PBD4 (**d**).

All samples were mixed with a linear RNA size standard prior to subjecting onto the gel. The first dimension gel of the respective system is implemented in each panel. The diagonal marks the linear RNAs; circular species are denoted by an arrow.

While identification of monomeric cleavage products was straight forward since they move roughly according to their size, identification of higher ligation products is challenging, because their movement can be irregular. Such atypical migration behavior of RNA in PAA-gels has been observed in the past (30), and suggests using RNA size markers rather as an approximate guide. However, by means of the 1-83mer marker (Fig. 4, lane 7) we know that the species moving around 150nt length is actually the circular 83mer, while the covalently linked linear 166mer (83+83) is located roughly at 120 nt length. Bands above the 200 bases ladder correspond most likely to 249mers (83+83+83) and even longer molecules. In comparison, we do see multiple species between 100 and 200nt in the full-length CRZ-2 lane (Fig. 4, lane 1). We can clearly identify the 166mer at the same height as the 166mer in the lane of the 1-83mer reference marker (lane 7). Shortly above is a stronger band indicating intermolecular ligation of a 83mer and a 92/94mer, respectively. The ratio between linear 166mer and 175/177mer would also be similar to the observed ratio between 83mer (**c** in lane 1) and 92/94mer (**b** in lane 1). The bands further up are hard to interpret and might show a little of c83mer and 186mer (92+94), as well as a 258/260mer (83+83+92/94) next to the 300bases ladder.

Assignment of bands becomes more difficult for PBD1 to 4. Since we cannot compare PBD1-4 to their linear 83mers, we can only hypothesize which multimers we observe. PBD1 (lane 3), our most efficient ribozyme concerning 5'- and 3'-end-cleavage, shows two bands in addition to the c83mer, which most likely correspond to linear 166mer (83+83) and 249mer (83+83+83), respectively. PBD2 (lane 4) shows four weak species between the 105mer (**a**) and the c83mer (**d**). Since we can see a clear band for 92/94mers (**b**) we suggest that these species took part in intermolecular ligation reactions with 83mers resulting in a diverse set of dimers (**e**). However, we cannot exclude that a low running trimer is present as well. PBD3 and PBD4 (lane 5, 6) show mostly the same species with different intensities. Analogous to PBD1 they clearly show two non-circular species that most likely represent the dimer (83+83) and trimer (83+83+83) respectively, both of them being located approximately at the same height as their CRZ-2 equivalents as shown by the l-83mer marker (lane 7).

Cyclic dimer formation

With the purpose of identifying cyclic dimers in the reaction mixture we designed and synthesized an inactive dimer (CRZ*) which should mimic the behavior of its CRZ-2 equivalent. Figure 6 shows two versions of non-reactive CRZ*, the linear species at roughly 166 nt length and the enzymatically ligated circular version at a height of roughly 800 bases (lane 1). By comparison with the results shown in Figure 4, a band at comparable height (~800 bases), is detected only in the l-83mer marker (lane 7), although being rather weak. All other ribozymes do not exhibit measurable amounts of circular dimers in PAA gels.

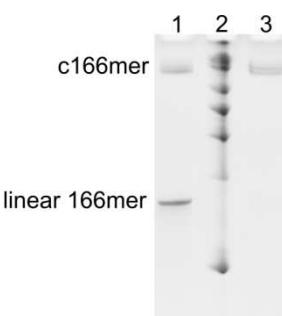


Figure 6: Enzymatic ligation of the inactive 166mer and treatment with exonuclease RNase R. Lane 1: ligation mixture composed of ligation product and remaining non-ligated linear transcript; lane 2: RNA size standard, 100 b, 200 b, 300 b, 400 b, 600 b, 800 b and 1000 b; lane 3: ligation product mixture after treatment with RNase R.

AFM measurements

AFM imaging is able to visualize RNA chains on single molecule level(31), allowing to characterize even rarely produced RNA species that are difficult (if not impossible) to observe in gel electrophoretic experiments. We analyzed four self-processing RNAs with AFM imaging under semi-denaturing conditions: CRZ-2, PBD1, PBD4 and the isolated linear 83mer of CRZ-2. These reaction mixtures showed high diversity upon biochemical analysis (Fig. 4 lanes 1, 3, 6, 7), with PBD1 forming predominantly cyclic 83mers, linear 166mers and 249mers, and PBD4 expressing a plethora of dimeric and of multimeric species.

Figures 7 and 8 show representative examples for tapping mode (TM) AFM images of ribozymes (recorded in air after RNA immobilization on mica and drying). The observed RNA chains adopt either a coiled (see white arrow in Fig. 7a), or uncoiled conformation which consists of rod-like segments, connected by kinks. Hence, for the uncoiled conformation it is possible to measure the lengths of the constituting segments as well as the contour length of the whole chain. That allows a comparison of the observed molecules with secondary structure prediction of the species listed in Table 2. Immobilized under native conditions, all observed molecules had a coiled conformation (data not shown), while semi-denaturing conditions resulted mostly in uncoiled conformations having the rod-kink-motif. Hence, the majority of the AFM measurements were done on RNA chains prepared under semi-denaturing conditions (see Figures 7 and 8 for a representative overview).

Histograms showing both the contour lengths and the segment lengths for all four analyzed ribozymes can be seen in Figures S4 and S5 (Supporting Information), contour length results are summarized in Table 3. These histograms are in agreement with the expected values from secondary and tertiary structure prediction: All single-molecule ribozyme species (83mer – 105mer) are expected to form a reactive structure with two stiff helical regions (segments) connected with a flexible kink. If we assume a typical pitch of 0.3 nm per base pair(31,32), the 83mer consists of two stiff regions with 5.4 and 6.3 nm length plus a kink of about five bases. The contour length would therefore be around 11.7 nm plus the kink region. Monomers that have non-cleaved ends would form the same helices but have additional single stranded regions in the kink region or sticking out from one of the helices. Based on these single stranded regions, different monomer variants would be hardly distinguishable with AFM imaging. Accordingly, different dimer species are expected to fold into a conformation where roughly 166 bases are involved in successive helical regions ($166 \times 0.15\text{nm} = 24.9\text{ nm}$), trimers with 249 bases resulting in 37.35 nm and so on.

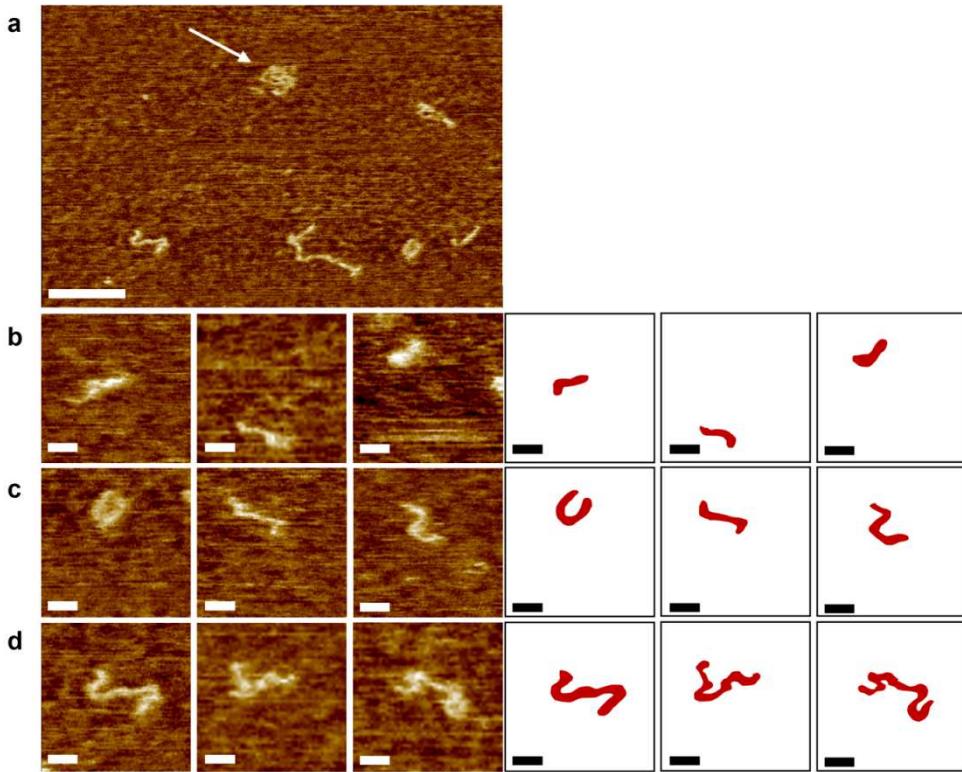


Figure 7: Tapping mode (TM) AFM phase images (range: 0 – 30°) of the reaction products resulting from incubation of the 1-83mer (isolated from CRZ-2 system) in cleavage/ligation buffer. For AFM analysis, samples were precipitated and resolved in 25 mM EDTA and 3.5 M urea(semi-denaturing conditions). Scale bars: 50 nm (**a**), 10 nm (**b-d**). The overview scan (**a**) shows RNA chains in coiled (white arrow) and unwrapped conformation. High resolution TM images (**b-d**) allow investigation of the internal structure of 83mers (**b**), dimers (**c**) and trimers (**d**). For convenience, schematics have been included on the right side to help with the interpretation of the AFM images.

The contour length histogram of the linear 83mer shows three contour length peaks at 13.4 nm, 24.5 nm and 36.5 nm, as well as very few molecules with even higher lengths (Fig.S5a). These peaks closely map to expected values for regularly folded monomers, dimers and trimers, respectively, making an interpretation straightforward. The segment lengths, representing individual helical regions, showed peaks at 5.9 nm, 8.5 nm and 13.8 nm (Fig. S4a), which most likely correspond to the monomer helices and dimer-variants of these helices. AFM imaging reveals that the monomer of the 1-83mer typically consists of one “short” and one “long” segment, which enclose an angle of roughly 110° (Fig. 7b). The dimer

may be composed of two, three or even four segments (Fig. 7c, left to right), while the trimer shows typically a very complicated internal structure (Fig. 7d). Hence, a large variety of possible conformations is observed for dimers and trimers in the AFM images.

Full-length CRZ-2, as well as PBD1 and PBD4 show an even wider spectrum of contour length peaks and chain conformations (see Figures S4 and 5), which is expected from the computational design and the biochemical analysis. In the AFM measurements, full-length CRZ-2 creates predominantly species being shorter than 24 nm (Fig. S5b). This ‘cut-off’ shifts to 36 nm for PBD1, while much longer RNA chains (up to ~80 nm) are observed for PBD4. Hence, the population shifts progressively to longer RNA products from CRZ-2 over PBD1 to PBD4, which is in agreement with the gel electrophoretic analysis.

All structures show contour length peaks at expected values close to those from the 1-83mer, which makes an identification of monomers, dimers and trimers straightforward. Monomers appear mostly in a rod-like conformation, and a kink (similar to the 1-83mer monomers) is rarely resolvable (Fig. 8, a3, a4, b3, b4, c3, c4). Dimers adopt L- and Z-like conformations (Fig. 8, a7, a8, b7, b8, c7, c8). Higher ligation products (trimers etc.) are currently only observed for PBD4, which can lead (similar to the 1-83mer) to very complicated and irregularly shaped internal chain structures (c9 to c14 in Fig. 8).

However, besides these species, also additional peaks are found at contour lengths that are (i) shorter than the expected value for a regularly folded monomer (6.5 nm and 9.0 nm, observed for CRZ-2, PBD1 and PBD4), (ii) between the monomer and dimer length (17.2 nm for CRZ-2, 19.0 nm for PBD1, and 20.0 nm for PBD4), or (iii) between the dimer and trimer length (31.7 nm for PBD1, 29.6 nm for PBD4). We can exclude that cleaved ends from processed full-length ribozymes would have a length of 6.8 nm or 9.5 nm in the AFM images (for the measurement conditions used in the experiments). Instead, we can show that the smaller peaks match very well to segment length measurements (Fig. S4 in the Supporting Informational Data), suggesting that only one of the two helices is resolved by AFM imaging. Accordingly, we know that the catalytically active structure involves tertiary interactions to closely orient both helices to each other. Uncleaved structures with single stranded regions in the kink region might favor the back folding of the helices despite semi-denaturing conditions, which are meant to destroy tertiary base pairs. The AFM images further support this interpretation: species having a contour length around 18.7 ± 1.4 nm (*i.e.*, between the monomer and dimer length) typically show an L-like conformation.

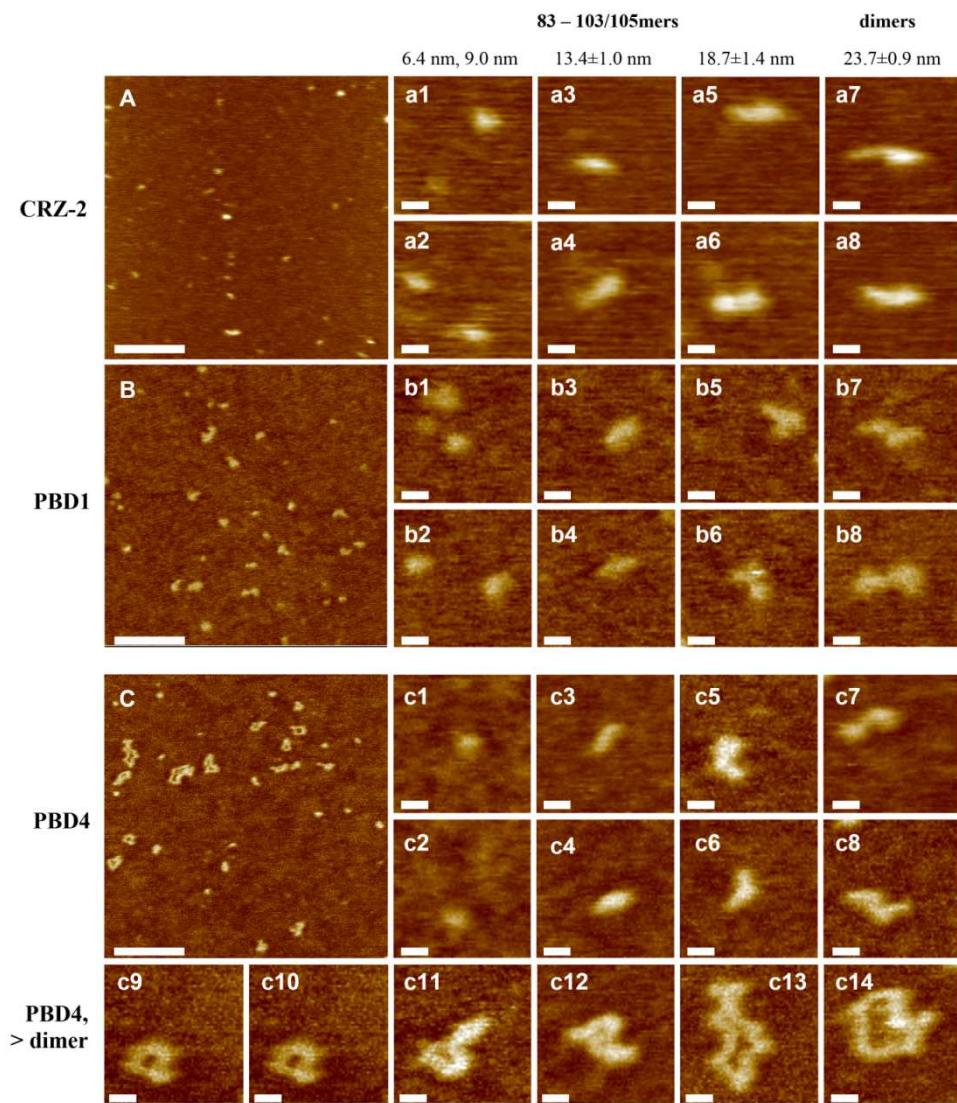


Figure 8: AFM images of RNA: CRZ-2 (**A**, **a1-a8**), PBD1 (**B**, **b1-b7**) and PBD4 (**C**, **c1-c14**). Scale bars: 100 nm (**A-C**), 10 nm (**a1-c14**); height scale 1 nm in all images. RNA chains have typically a height of ~0.4 nm in the AFM images. Overview scans (**A-C**) show a mixture of RNA chains of different contour length L_C for all three sequences investigated. Analysis of contour length histograms (see Figures S5 and S6 in the Supporting Information) for CRZ-2, PBD1 and PBD4 allowed association of most of the observed RNA chains with the species listed in Tab. 2 (as indicated in the figure). This procedure failed for two shortest species (**a-c 1,2**) and the one having a contour length around 18.7 nm (**a-c3,4**), as the calculated numbers of bases did not match any entry of this table (see text for a detailed discussion).

Complementing this chain structure with a third segment (which might be irresolvable in the images due to back folding of one helix) having a length of 6.6 ± 0.4 nm (first peak in the segment length histograms, see Fig. S5) gives a Z-like conformation with a total contour length of 25.3 ± 1.8 nm, matching well the expected value for a regularly folded dimer (24.9 nm). Using the same reasoning, the peaks around 6.4 ± 0.3 nm and 9.0 ± 0.6 nm may be interpreted as ‘partially back folded’ monomer and the one around 30.7 ± 1.5 nm as a ‘partially back folded’ trimer.

To compare the AFM findings with the computational design and the results of the biochemical analysis, we calculated the number frequency of each species from the contour length histograms. The ratio between observed monomers, dimers and trimers are given in the last two lines of Table 3: one line calculates the ratios regarding all peaks and one line regards only those peaks in the contour length histograms that exactly matched the expected contour lengths. However, both lines show very similar ratios, indicating that using the ‘backfolding hypothesis’ does not affect the final conclusions of the AFM measurements.

Table 3:AFM contour length measurements and their implications on the ratio between monomers (M), dimers (D) and trimers (T). Bold values correspond to contour length exactly matching expected values.

Species length number of bases	AFM contour length measurements (nm)			
	CRZ-2 1-83mer	CRZ-2	PBD1	PBD4
M (83 – 103/105)	13.4	6.8, 9.5, 12.3	6.2, 9.3, 13.8	6.3, 8.3, 14.2
D (166 – 186/188)	24.5	17.2, 22.7	19.0, 24.5	20.0, 24.0
T (249 – 269/271)	36.5		31.7, 36.5	29.6, 34.8
Ratio (M : D : T)	2 : 1.5 : 1	7.4 : 1 : 0	10 : 2.8 : 1	4.5 : 1 : 1
Bold Ratio only (M : D : T)	2 : 1.5 : 1	6 : 1 : 0	9.5 : 4 : 1	2.5 : 1 : 1

DISCUSSION

Taken together the results of the biochemical analysis in combination with AFM imaging confirm the predicted behavior of the self-processing RNAs CRZ-2, PBD1, PBD2, PBD3 and PBD4. All RNAs undergo two initial cleavage events that truncate the full length transcript at the 5'- and 3'-end to a linear 83mer with 5'-hydroxyl group and 2',3'-cyclic phosphate required for ligation. The subsequent intramolecular ligation delivers exclusively cyclic versions of the

83mer, whereas intermolecular ligation produces dimers and longer concatemers, which apart from PBD4, have no or rather low tendency towards cyclization. This implies that formation of cyclic dimers is extremely unfavored, since it requires the ligation at two sites simultaneously. The same applies to longer concatemers that are rather rare anyway.

Comparing the outcome of the experimental studies with what was theoretically predicted for the designed sequences, we note that there is compliance as well as contradiction as discussed below. Figure 9 shows a detailed analysis for each ribozyme and will serve as a guideline to discuss observed results from PAA gel electrophoresis and AFM. During the cleavage cascade, we can distinguish three types of reaction steps: (i) formation of reactive structures, (ii) dissociation of cleaved ends after ribozyme reaction and (iii) refolding of an unbound reaction product into a new reactive structure. In Figure 9 each of these steps is characterized by an activation free energy (see Supporting Information for details).

*Full length transcripts (**a** – 103/105mers)*

Our theoretical analysis shows that dissociation of the cleaved ends from computationally optimized ribozymes (PBD1-4) has to overcome a higher energy barrier than in the case of manually designed CRZ-2 (Fig. 9). This is due to the fact that designed molecules are optimized to fold primarily into catalytically active conformations, and therefore also the cleaved conformations with tightly bound ends are very stable. It is known that tightly bound fragments shift equilibrium towards ligation (16), and this would explain why we see full-length product for three of our four designed ribozyme species (PBD2, PBD3 and PBD4), but not for CRZ-2. Here, the 5'-end is efficiently cleaved off, and the resulting transient fragments tend to accumulate as 92mers (Fig. 4). The full length transcript PBD1 (103mer) is completely consumed despite its high dissociation barriers, indicating that in agreement with theoretical analysis, the dissociation of cleaved ends is an irreversible step at experimental conditions.

*Cleavage products (**b** – 97, 94, 92, 91mer and **c** - linear 83mer)*

The cleavage cascade can start with either of two reactive conformations resulting in cleavage of the 5'- or 3'-end. In case of CRZ-2, none of these conformations correspond to the ground state of the molecule, rather they are 4.2 and 8.5 kcal/mol above the ensemble free energy. Cleavage of the 5'-end is the favored reaction, but results in a structure that has to overcome a high barrier to fold into the reactive 92mer conformation for 3'-end cleavage.

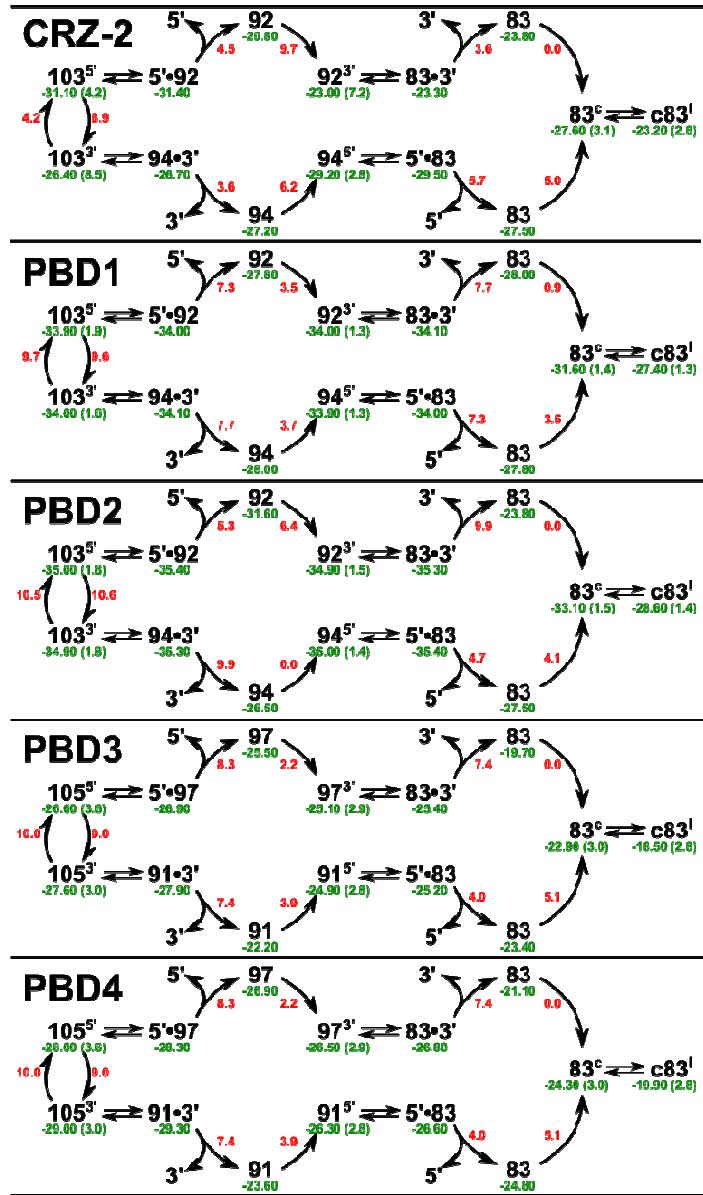


Figure 9: Cleavage cascades of molecules CRZ-2 and PBD1-4. Black numbers correspond to the length of the molecules or to the fragment to be cleaved (5'- and 3'-end). Superscripts 5', 3', c or l mark molecules in a reactive conformation to cleave the 5'- end, the 3'-end, to circularize or to linearize, respectively. Reversible cleavage reactions are indicated by double arrows, bent arcs denote refolding steps that are considered as irreversible. Green numbers show the free energies of the respective molecules. A green number in parentheses denotes the difference between the free energy and the ensemble free energy, which determines the equilibrium probability of the reactive conformation. Red numbers show the free energy barriers that need to be overcome in order to refold one conformation into the next conformation.

In equilibrium, the reactive structure is sparsely populated, since it lies 7.2 kcal/mol above the ensemble free energy. We therefore expect that the prominent band **b** in the CRZ-2 lane in Figure 4 is mostly 92mer, since the 94mer is (*i*) the less favored cleavage product and (*ii*) more likely to undergo the following-up cleavage reaction.

PBD1 and PBD2 are the highly optimized sequences for active 83mer production, and they show very similar high probabilities to fold into both catalytically active conformations (see Figures 2 and 3) in equilibrium. However, looking at dissociation and refolding barriers, PBD1 has two equally good pathways, in contrast to PBD2, which favors cleavage of the 5'-end. We would therefore expect that PBD2 bands **a** and **b** in Figure 4 (lane 4) correspond to the 103mer and 92mer, prone to fold in the conformation required to cleave the 3'-end.

PBD3 and PBD4 enable a clear separation of cleavage products on the gel picture, due to their differently sized ends. Both molecules differ by only two point mutations in one hairpin loop of the reactive conformations (Fig. 3). Since this hairpin remains closed in all reactive species as well as on the most favorable refolding paths between the species, all important free energies differ by a constant factor (1.4 kcal/mol), and the barrier heights and structure ensemble probabilities are the same. The distribution of monomeric species should therefore be exactly the same for PBD3 and PBD4. Both molecules favor to cleave first the 3'-end, and second the 5'-end. This is in accordance with experimental results, showing mainly the 91mer as favored intermediate product. The 97mer is seen, particularly for PBD4, when higher amounts of RNA are subjected onto the gel (Supporting Information, Fig. S3d).

*Higher non-cyclic ligation products (**e** – dimers, trimers, concatemers)*

All hairpin-ribozyme variants can form two long helices, both of which have the possibility to form stable dimers that preserve the feature of catalytic activity. PBD1 and PBD3 have self-complementary hairpin loops, which results in a generally stronger tendency to dimerize, and a higher probability to retain catalytic activity upon dimerization. Since PBD1 is extremely efficient during the monomeric cleavage cascade, it shows only minor amounts of intermediate cleavage products (Supporting Information, Fig. S3a) that could form dimers. Accordingly, the only dimer species we see in Figure 4 is the 166mer (**e** in lane 3). CRZ-2 and PBD2 show the greatest variety of concatemeric species. In case of CRZ-2, we see stable 92-/94- and 83mer cleavage products (Fig. 4, bands **b** and **c** in lane 1). The probabilities to ligate their reactive ends are higher than the probabilities to cleave off the remaining terminal sequence patches, which corresponds to the fact that we observe a variety of

concatemeric species. PBD2 shows only the 92-/94mer band **b** and no band for the 83mer **c** (Fig. 4, lane 4), but it has the highest probabilities (after PBD1) to ligate intermediate cleavage products, and, in contrast to PBD1, a low probability to cleave reactive ends upon dimerization.

PBD3 and PBD4 differ only in the self-complementarity of one hairpin loop and thus should be the best systems to study the influence of such mutations. Resulting from perfect self-complementarity, PBD3 has both higher probabilities to ligate intermediate products and higher probabilities to cleave ends from dimer species. However, the only detectable cleavage intermediate on gel pictures for both PBD3 and PBD4 is the 91mer, which cannot ligate with itself to a higher species. Accordingly, we see exclusively the linear 166mer (dimer) for both species. Interestingly, PBD4 shows more multimeric species, suggesting that design towards stable dimers (PBD3) leads to a lower diversity of multimers.

AFM visualization of RNA molecules

AFM imaging is able to resolve the shape of RNA molecules, even short ribozymes (< 100 bases). RNA species are identified using differences in their contour length, which can cause ambiguities if species differ only by a few nanometers. Hence, it was not possible to distinguish linear and cyclic species (same contour length), different monomeric cleavage products from the full length transcript (contour lengths differ by less than 2 nm), or to distinguish, which type of dimer, trimer, etc. is observed in the AFM image. However, AFM resolved structural features (helices, loop regions) of different species and observed segment and contour lengths that match with secondary structure prediction for monomers, dimers and trimers. In case of the linear 83mer of CRZ-2 (which can only form multiples of itself) the typical pitch of 0.30 ± 0.02 nm, per base pair in a helix(31,32) matches exactly our observed segment and contour lengths. Contour lengths that do not match the values of regularly folded monomers, dimers and trimers can be explained by spatial proximity of two adjacent segments, such that two segments appear as one. Supporting this hypothesis, adding a single segment from the segment length histogram to truncated RNA species would lead to expected contour lengths.

We furthermore observed that, although the samples were dried before imaging, the RNA chains kept most of the initial helical conformation. This was also observed in earlier studies (31,33,34), and indicates that the RNA chain structure is sufficiently conserved to yield meaningful results using AFM imaging in air. Tip convolution, which may lead to a systematic overestimation of the contour lengths, introduces only minor disturbances. Using

typical experimental parameters (tip radius < 5 nm, RNA chain height < 0.4 nm), tip convolution increases the lateral chain extension by less than 3 nm (measured as full width half maximum/FWHM) (35), which is less than 20% of the monomeric contour length, less than 10% of the dimeric one, etc. However, the good quantitative agreement suggests that tip convolution effects, (*i.e.*, the effective tip radius) are smaller than expected for tip radii extracted from calibration measurements as described in the Experimental section. Taken together, the results of AFM measurements confirmed and complemented the conclusions drawn from the gel electrophoretic analysis. While smaller fragments are dominating for CRZ-2, a tendency towards larger constructs is seen for PBD1, and for PBD4 a majority of rather complex structures is detected (Fig. 8). Comparing the outcome of the AFM analysis for PBD4 with the gel shown in Figure 4, these complex structures are either cyclic species of varying lengths or folded concatemers, since they most likely correspond to the smear of bands in the area of d (Fig. 4, lane 4, comp. also Fig. 5d and Fig. S3d), and thus show an anomalous migration behavior as typically observed for cyclic or folded RNA species(21,29). We obtain clear results from contour length measurements counting *relative* amounts of monomers, dimers and trimers (Table 3). Regardless of whether we compare ratios of theoretically expected peaks only, or include the peaks corresponding to partly unresolved molecules, we observe more dimeric and trimeric species for our newly designed species PBD1 and PBD4, which is in agreement with our design objective. Furthermore, PBD1 tends to form dimer species, again in agreement with our design goal, while PBD4, which was theoretically optimized to form cyclic monomers, shows the highest multimer variety both on PAA gel electrophoresis and AFM imaging.

To summarize, imaging ribozymes on the single-molecule level using AFM provides information that complements results obtained from the gel electrophoretic analysis and the computation analysis (and *vice versa*), making a combination of these techniques promising and very powerful.

CONCLUSION

Through the combination of computational prediction, biochemical and biophysical analysis, our study revealed that:(*i*)self-processing activity canbe programmed into RNA structures, (*ii*)self-processing activity can be predicted and optimized by computer-aided design, and (*iii*) AFM turned out to be a powerful technique to image the reaction products at the single molecule level, even for short RNAs (<100mer).

In silico prediction was highly effective for the design of our self-processing RNAs, while gel electrophoretic analysis complemented by AFM imaging has proven very suitable for reaction analysis. In particular 2D gel electrophoresis allowed us to differentiate between cyclic and linear species, whilst AFM delivered information on RNA lengths, enabling us to determine the distribution of the various RNA products in individual reaction mixtures. Our combined approach will have a strong impact on the field of RNA engineering for solving complex design problems. Dynamic processes like self-induced topology changes and oligomerization and their sensitivity upon sequence variation are essential for biological self-organization and evolution. Moreover, a large number of publications over the past two years have shown that biological processing of RNA into circular species with often still unknown function is widely spread in nature. Thus, nowadays the emergence of circular RNAs and their cellular functionalities are actively investigated (36-40), making the development of *in vitro* techniques for RNA circularization and the study of models mimicking the processing into circular species even more important.

ACKNOWLEDGMENT

S.B. and M.D. acknowledge support by the German Federal Ministry of Education and Research of Germany (BMBF) within the project FKZ 03Z2CN11.I.H. and S.M. acknowledge support by FWF (Austria) and DFG (Germany) within the program ERA chemistry.

REFERENCES

1. Flores, R., Grubb, D., Elleuch, A., Nohales, M.A., Delgado, S. and Gago, S. (2011) Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. *RNA Biol.*, **8**, 200-206.
2. Hampel, A. and Tritz, R. (1989) RNA catalytic properties of the minimum (-)sTRSV sequence. *Biochemistry*, **28**, 4929-4933.
3. DeYoung, M., Siwkowski, A.M., Lian, Y. and Hampel, A. (1995) Catalytic properties of hairpin ribozymes derived from Chicory yellow mottle virus and arabis mosaic virus satellite RNAs. *Biochemistry*, **34**, 15785-15791.
4. Hammann, C., Luptak, A., Perreault, J. and de la Pena, M. (2012) The ubiquitous hammerhead ribozyme. *RNA*, **18**, 871-885.
5. Muller, S., Appel, B., Krellenberg, T. and Petkovic, S. (2012) The many faces of the hairpin ribozyme: structural and functional variants of a small catalytic RNA. *IUBMB Life*, **64**, 36-47.
6. Petkovic, S. and Muller, S. (2013) RNA self-processing: Formation of cyclic species and concatemers from a small engineered RNA. *FEBS Lett.*, **587**, 2435-2440.
7. Drude, I., Strahl, A., Galla, D., Muller, O. and Muller, S. (2011) Design of hairpin ribozyme variants with improved activity for poorly processed substrates. *FEBS J.*, **278**, 622-633.

8. Drude, I., Vauleon, S. and Muller, S. (2007) Twin ribozyme mediated removal of nucleotides from an internal RNA site. *Biochem Biophys Res Commun*, **363**, 24-29.
9. Strohbach, D., Novak, N. and Muller, S. (2006) Redox-active riboswitching: allosteric regulation of ribozyme activity by ligand-shape control. *Angew Chem Int Ed Engl*, **45**, 2127-2129.
10. Vauleon, S., Ivanov, S.A., Gwiazda, S. and Muller, S. (2005) Site-specific fluorescent and affinity labelling of RNA by using a small engineered twin ribozyme. *Chembiochem*, **6**, 2158-2162.
11. Ivanov, S.A., Vauleon, S. and Muller, S. (2005) Efficient RNA ligation by reverse-joined hairpin ribozymes and engineering of twin ribozymes consisting of conventional and reverse-joined hairpin ribozyme units. *FEBS J*, **272**, 4464-4474.
12. Welz, R., Bossmann, K., Klug, C., Schmidt, C., Fritz, H.J. and Muller, S. (2003) Site-directed alteration of RNA sequence mediated by an engineered twin ribozyme. *Angew Chem Int Ed Engl*, **42**, 2424-2427.
13. Vauleon, S. and Muller, S. (2003) External regulation of hairpin ribozyme activity by an oligonucleotide effector. *Chembiochem*, **4**, 220-224.
14. Pieper, S., Vauleon, S. and Muller, S. (2007) RNA self-processing towards changed topology and sequence oligomerization. *Biol Chem*, **388**, 743-746.
15. Cochrane, J.C. and Strobel, S.A. (2008) Catalytic strategies of self-cleaving ribozymes. *Acc Chem Res*, **41**, 1027-1035.
16. Fedor, M.J. (1999) Tertiary structure stabilization promotes hairpin ribozyme ligation. *Biochemistry*, **38**, 11040-11050.
17. McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105-1119.
18. Lorenz, R., Bernhart, S.H., Honer zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**, 26.
19. Bernhart, S.H., Tafer, H., Muckstein, U., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol*, **1**, 3.
20. Harris, M.E. and Christian, E.L. (1999) Use of circular permutation and end modification to position photoaffinity probes for analysis of RNA structure. *Methods*, **18**, 51-59.
21. Sigurdsson, S.T. and Eckstein, F. (1996) Isolation of oligoribonucleotides containing intramolecular cross-links. *Anal Biochem*, **235**, 241-242.
22. Umekage, S. and Kikuchi, Y. (2009) In vitro and in vivo production and purification of circular RNA aptamer. *J Biotechnol*, **139**, 265-272.
23. Pasman, Z., Been, M.D. and Garcia-Blanco, M.A. (1996) Exon circularization in mammalian nuclear extracts. *RNA*, **2**, 603-610.
24. Rivetti, C., Guthold, M. and Bustamante, C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J Mol Biol*, **264**, 919-932.
25. Rivetti, C. and Codeluppi, S. (2001) Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B- to A-form transition on mica. *Ultramicroscopy*, **87**, 55-66.
26. Berzal-Herranz, A., Joseph, S., Chowrira, B.M., Butcher, S.E. and Burke, J.M. (1993) Essential nucleotide sequences and secondary structure elements of the hairpin ribozyme. *EMBO J*, **12**, 2567-2573.
27. Flamm, C., Hofacker, I.L., Maurer-Stroh, S., Stadler, P.F. and Zehl, M. (2001) Design of multistable RNA molecules. *RNA*, **7**, 254-265.

28. Butcher, S.E. and Burke, J.M. (1994) A photo-cross-linkable tertiary structure motif found in functionally distinct RNA molecules is essential for catalytic function of the hairpin ribozyme. *Biochemistry*, **33**, 992-999.
29. Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1984) Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell*, **37**, 415-427.
30. Cruz-Reyes, J., Piller, K.J., Rusche, L.N., Mukherjee, M. and Sollner-Webb, B. (1998) Unexpected electrophoretic migration of RNA with different 3' termini causes a RNA sizing ambiguity that can be resolved using nuclease P1-generated sequencing ladders. *Biochemistry*, **37**, 6059-6064.
31. Henn, A., Medalia, O., Shi, S.P., Steinberg, M., Franceschi, F. and Sagi, I. (2001) Visualization of unwinding activity of duplex RNA by DbpA, a DEAD box helicase, at single-molecule resolution by atomic force microscopy. *Proc Natl Acad Sci U S A*, **98**, 5007-5012.
32. Arnott, S., Hukins, D.W.L., Dover, S.D., Fuller, W. and Hodgson, A.R. (1973) STRUCTURES OF SYNTHETIC POLYNUCLEOTIDES IN A-RNA AND A'-RNA CONFORMATIONS - X-RAY-DIFFRACTION ANALYSES OF MOLECULAR-CONFORMATIONS OF POLYADENYLIC-ACID POLYURIDYLIC ACID AND POLYINOSINIC ACID POLYCYTIDYLIC ACID. *Journal of Molecular Biology*, **81**, 107-&.
33. Abels, J.A., Moreno-Herrero, F., van der Heijden, T., Dekker, C. and Dekker, N.H. (2005) Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys J*, **88**, 2737-2744.
34. Bonin, M., Oberstrass, J., Lukacs, N., Ewert, K., Oesterschulze, E., Kassing, R. and Nellen, W. (2000) Determination of preferential binding sites for anti-dsRNA antibodies on double-stranded RNA by scanning force microscopy. *RNA*, **6**, 563-570.
35. Ortinau, S., Schmich, J., Block, S., Liedmann, A., Jonas, L., Weiss, D.G., Helm, C.A., Rolfs, A. and Frech, M.J. (2010) Effect of 3D-scaffold formation on differentiation and survival in human neural progenitor cells. *Biomed Eng Online*, **9**, 70.
36. Danan, M., Schwartz, S., Edelheit, S. and Sorek, R. (2012) Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res*, **40**, 3131-3142.
37. Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. and Sharpless, N.E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, **19**, 141-157.
38. Liu, Y., Cui, H., Wang, W., Li, L., Wang, Z., Yang, S. and Zhang, X. (2013) Construction of circular miRNA sponges targeting miR-21 or miR-221 and demonstration of their excellent anticancer effects on malignant melanoma cells. *Int J Biochem Cell Biol*, **45**, 2643-2650.
39. Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J. and Kjems, J. (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J*, **30**, 4414-4422.
40. Abe, N., Abe, H. and Ito, Y. (2012) Synthesis of dumbbell-shaped cyclic RNAs for RNA interference. *Curr Protoc Nucleic Acid Chem*, **Chapter 16**, Unit 16 14 11-11.

Supporting Information:

Sequence-controlled RNA self-processing: computational design, biochemical analysis and visualization by AFM

Sonja Petkovic, Stefan Badelt, Stephan Block, Christoph Flamm, Mihaela Delcea, Ivo Hofacker, Sabine Müller

Corresponding authors:

Prof. Dr. Sabine Müller; Email: smueller@uni-greifswald.de

Prof. Dr. Ivo Hofacker; Email: ivo@tbi.univie.ac.at

Dr. Stephan Block; Email: stephan.block@chalmers.se

Table of Contents

Figure S1: Models of the monomer and dimer cleavage cascades	SI 2
Figure S2: Products of ribozyme reactions in preparative scale	SI 3
Figure S3: 2D-Analyses of reference systems	SI 4
Figure S4: AFM segment length histograms of reference and test systems	SI 6
Figure S5: AFM contour histograms of references and test systems PBD1 and 4	SI 7
Estimation of activation energies (Figure 9)	SI 8
Table S1: Klenow primer sequences and References	SI 9

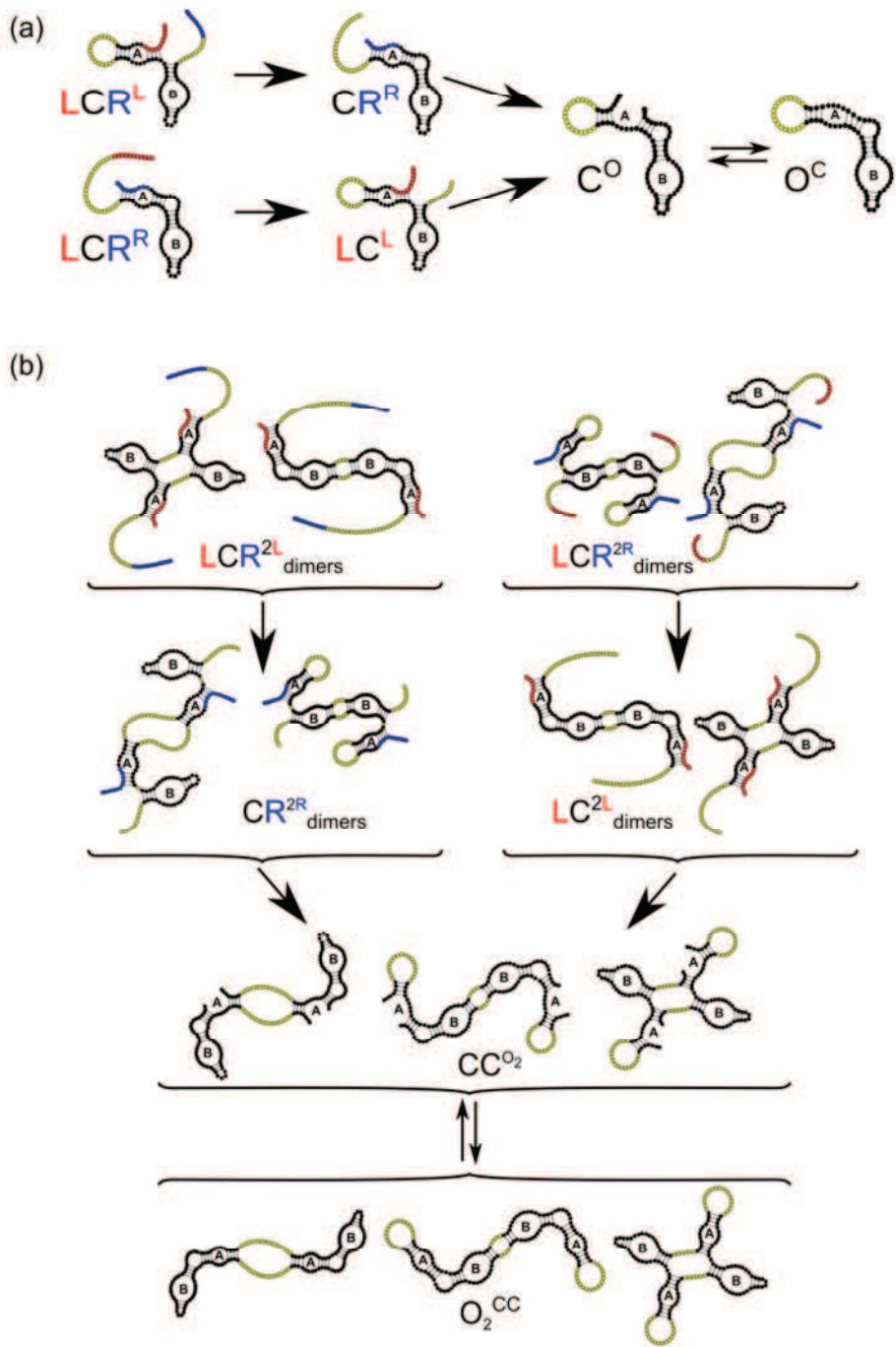
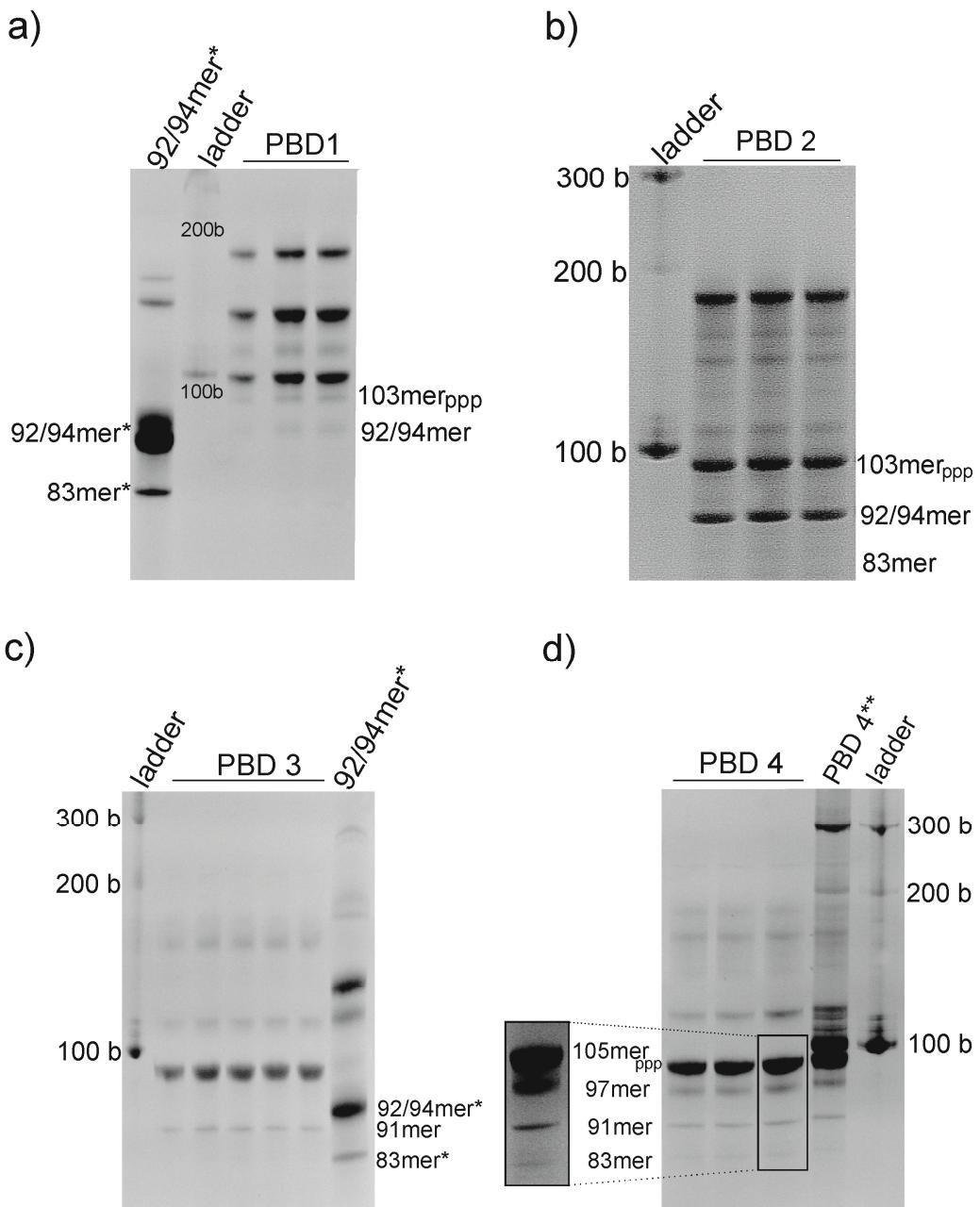


Figure S1: Models of (a) the monomer and (b) dimer cleavage cascade

Red and blue colored regions mark cleavable 5'- and 3'-ends, respectively. A cleavage/ligation reaction can only occur when tertiary interactions between loop A and loop B are formed. In black we show structure constraints needed for such reactions, while yellow regions should be flexible without impairing catalytic activity. Importantly, every structure constraint defines a non-overlapping set of structures such that the probability of forming a reactive molecule can be computed from the sum of the constraint partition functions. RNA secondary structures were drawn using jViz (1).



*origin from CRZ-2-RNA

** mixture of PBD 4 and ladder

Figure S2: Self-processing products of the four designed RNAs PBD1 to 4 analyzed with a 15% denaturing polyacrylamide gel (preparative scale).

- a:** PBD1 self-processing products upon reaction at PBD1 starting concentration of 2.5 μ M;
- b:** PBD2self-processing products upon reaction at PBD2 starting concentration of 3 μ M;
- c:** PBD3self-processing products upon reaction at PBD3 starting concentration of 2 μ M;
- d:** PBD4self-processing products upon reaction at PBD4 starting concentration of 2 μ M. For better visualization, the boxed area is shown slightly magnified and at higher contrast on the left.

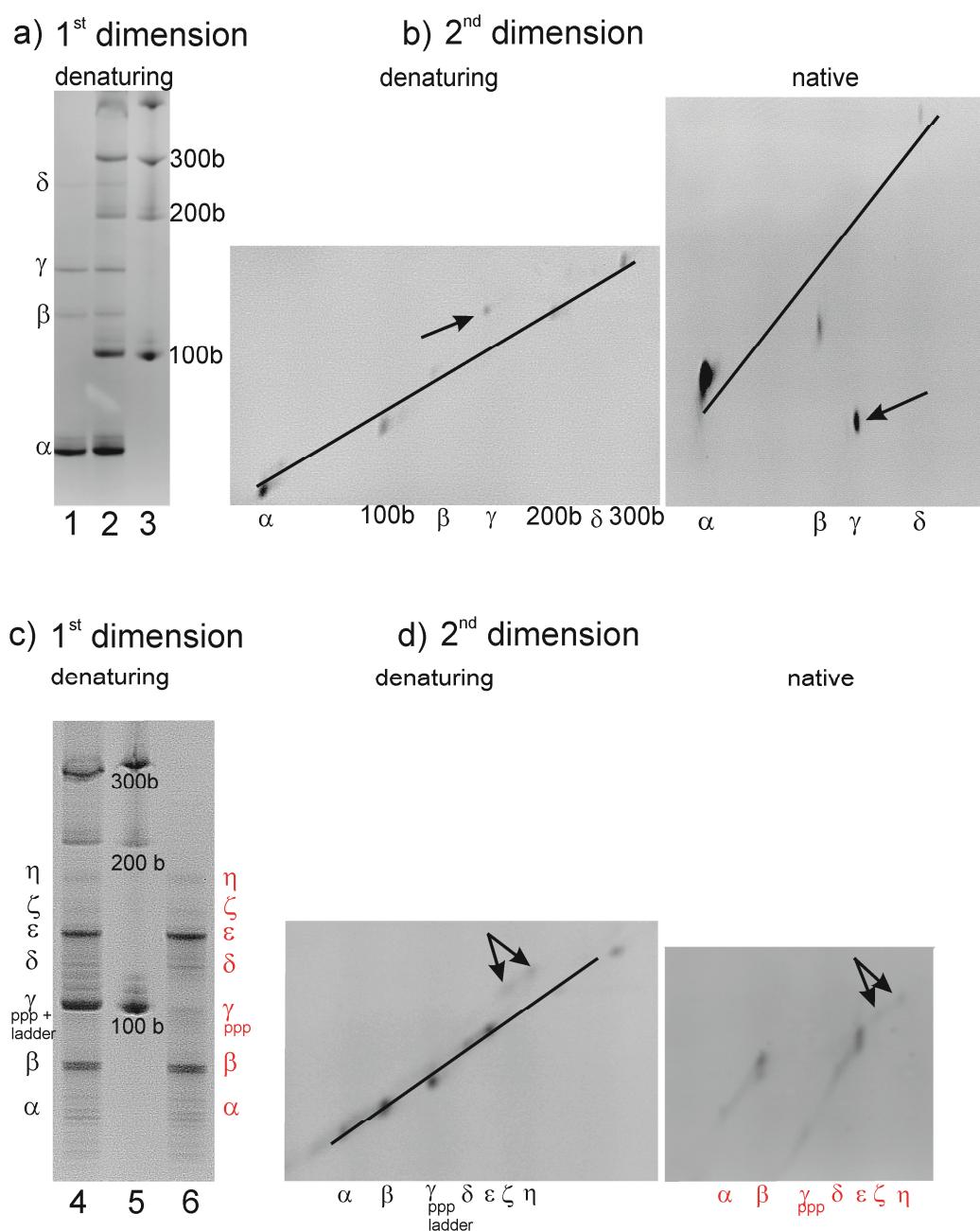


Figure S3: Identification of cyclic RNA from I-83mer and from full-length CRZ-2 by 2D-PAGE.

a: Self-processing of I-83mer analyzed on a 15%denaturing polyacrylamide gel. Lane 1: self-processing productsdenoted with Greek lettersα to δ. Lane 2: self-processing products mixed with linear RNA size standard. Lane 3: linear RNA size standard. **b:** Second-dimension denaturing (left, to improve resolution polyacrylamide concentration was increased to 17.5%) and native (right, 15%) polyacrylamide gels. Lane 1 of the gel shown in panel (a) was cut off and used as "starting slot" for the native gel in second dimension (b, right), lane 2 of the gel shown in panel (a) was equally used for the denaturing gel in second dimension (b,

left). Species γ (marked by an arrow) appears beyond the diagonal in both gels, implying its cyclic nature. **c:** Self-processing of full length CRZ-2 (103mer) analyzed on a 15% denaturing polyacrylamide gel. Lane 4: self-processing products denoted with Greek letters α to η , mixed with linear RNA size standard. Lane 5: linear RNA size standard. Lane 6: self-processing products denoted with Greek letters α to η . **d:** Second-dimension denaturing (left, to improve resolution polyacrylamide concentration was increased to 17.5%) and native (right, 15%) polyacrylamide gels. Lane 4 of the gel shown in panel **(c)** was cut off and used as "starting slot" for the denaturing gel in second dimension (**d**, left), lane 6 of the gel shown in panel **(c)** was equally used for the native gel in second dimension (**d**, right). The two blurry spots ζ and η marked by arrows in the denaturing gel in panel **(d)** might correspond to cyclic RNAs. However, the corresponding native gel on the right reveals that both RNAs do not migrate as fast as would be expected for cyclic species according to the analysis of I-83mer shown in panel **(b)**.

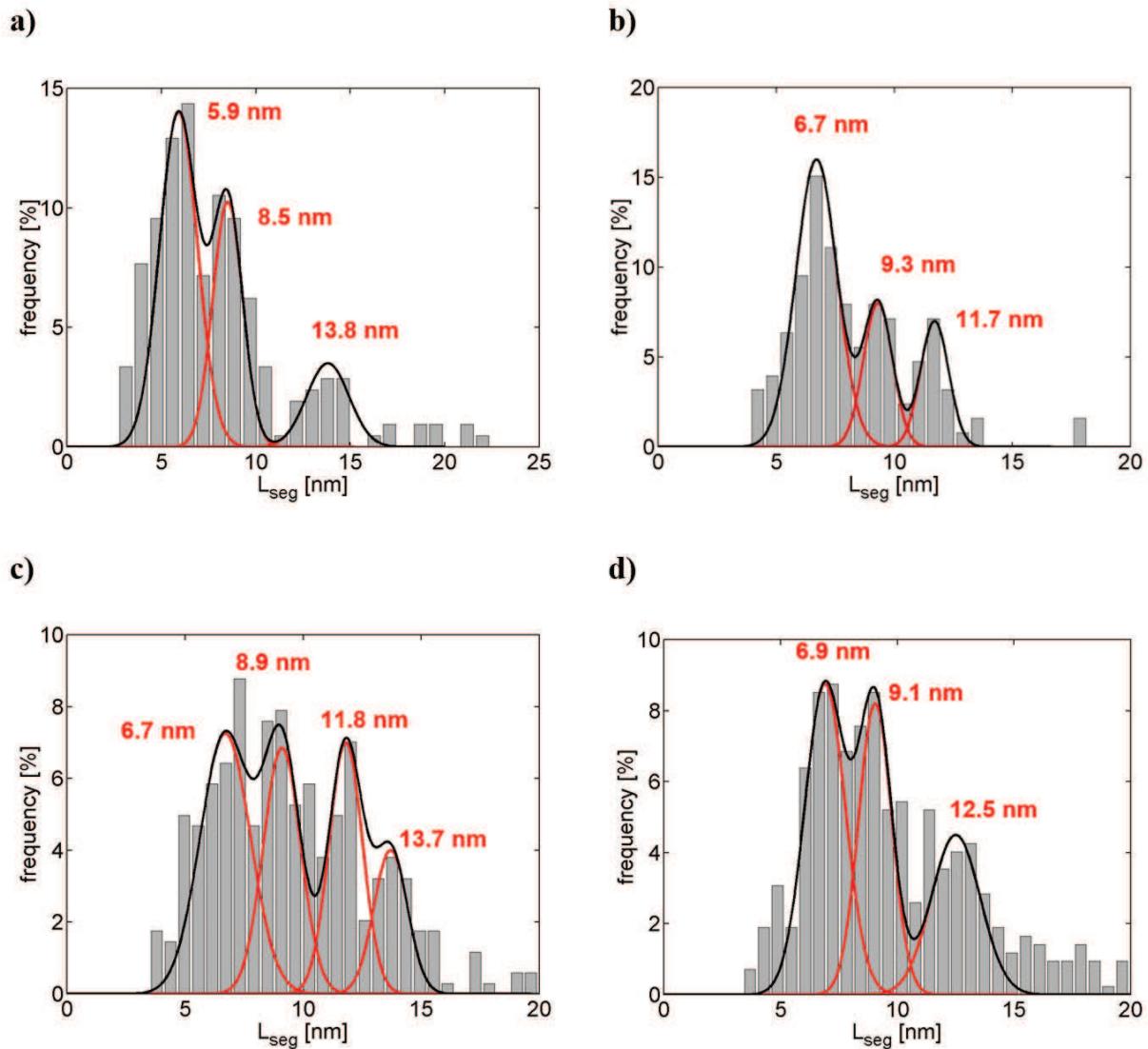


Figure S4: RNA segment length histograms of references: (a) I-83mer, and (b) CRZ-2and test sequences (c) PBD1 and (d) PBD4

In tapping mode AFM images the RNA chains typically consist of rod-like segments, which are connected via kinks (see Figs 5 and 6). From the AFM images, the length of these segments can be measured with nm accuracy, showing several well-resolvable peaks at (average \pm standard deviation as averaged over the 4 different RNA sequences) 6.6 ± 0.4 nm, 9.0 ± 0.3 nm, 12.0 ± 0.4 nm, and 13.8 ± 0.1 nm. Assuming an alpha helical conformation of the rod-like segments and therefore a typical pitch of 0.3 nm per base pair, these peaks correspond to segments consisting of 22 ± 1 bp, 30 ± 1 bp, 40 ± 1 bp, 46 ± 0.3 bp. These histograms include the length of (a) 212, (b) 127, (c) 342 and (d) 423 segments.

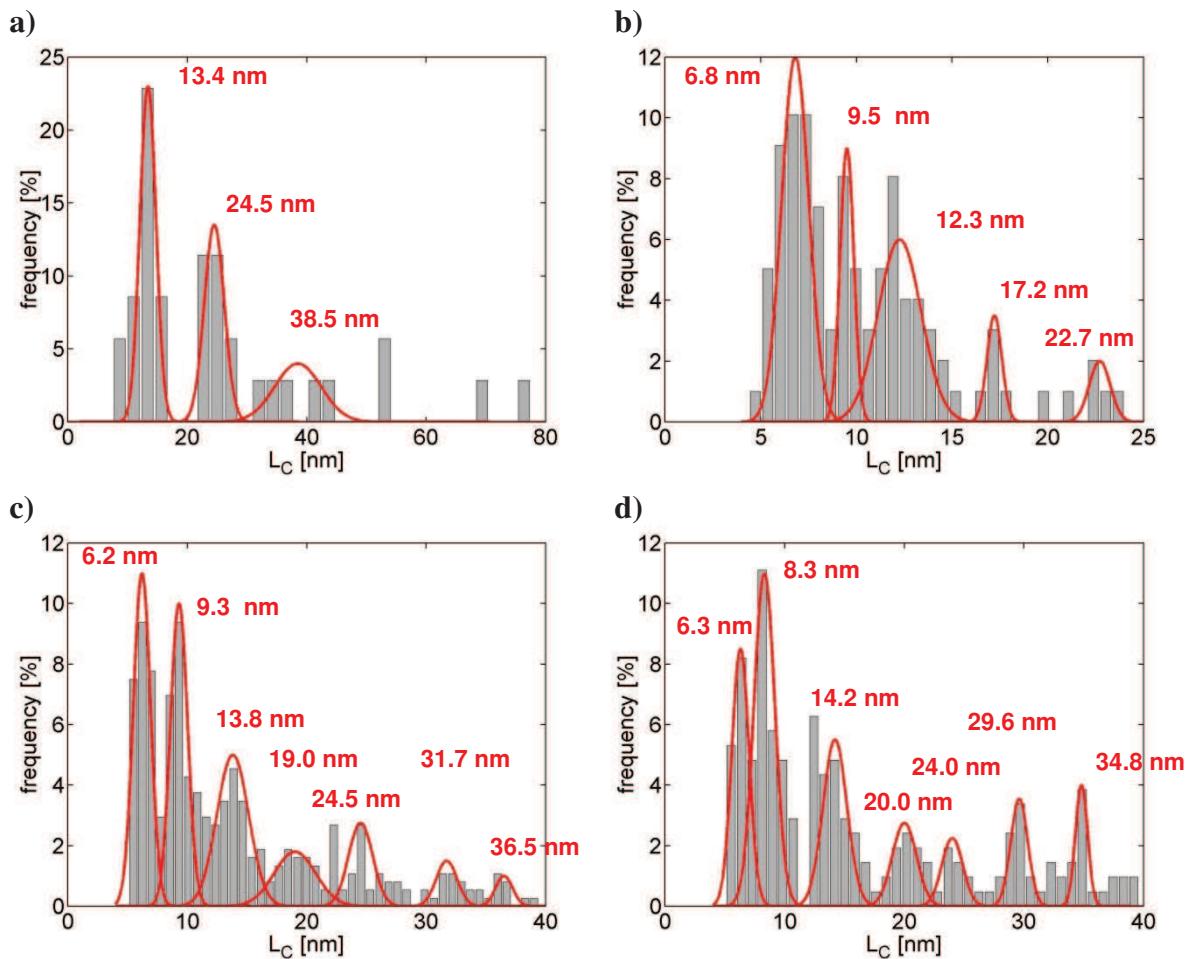


Figure S5: Contour length histograms of RNA chains for (a) I-83mer, (b) CRZ-2, (c) PBD1, and (d) PBD4.

For the linear 83mer, the histogram gives three peaks at 13.4, 24.5, and 36.5 nm. Assuming again an alpha helical conformation and therefore a typical pitch of 0.3 nm per base pair (= 0.15 nm per base), these peaks correspond to 89, 163 and 243 bases and can be identified as final cleavage products (83mer) and higher ligation products (dimer = 166mer; trimer = 249mer). (b-d) As the other RNA constructs additionally create intermediate cleavage products, their contour length histograms exhibit a more complicated peak structure. Generally, peaks are observed in all histograms at similar values allowing averaging the determined peak position over the three different RNA sequence. This yield the following values (average \pm standard deviation of the respective peak position): 6.4 \pm 0.3 nm (first peak), 9.0 \pm 0.6 nm (second peak), 13.4 \pm 1.0 nm (third peak), 18.7 \pm 1.4 nm (fourth peak), 23.7 \pm 0.9 nm (fifth peak), and for PBD1 and PBD4 at 30.7 \pm 1.5 nm (sixth peak), and 35.7 \pm 1.2 nm (seventh peak). Note that the first two peaks coincide with the first two peaks of the segment length histograms in Figure S5. Hence, it is very likely that these peaks correspond to RNA chains, for which only one of the constituting segments was resolvable in the AFM image (e.g., if two neighbouring segments enclose an angle of approximately 180° and therefore appear as a single segment in the measurement). These histograms include the contour length of (a) 73, (b) 100, (c) 218 and (d) 256 RNA chains.

Estimation of activation energies (Figure 7)

Figure 7 in the main paper shows a comprehensive view of the cleavage cascade for each of the experimentally tested ribozymes. We can distinguish three types of reaction steps (*i*) formation of reactive structures, (*ii*) dissociation of cleaved ends after ribozyme reaction, (*iii*) refolding of an unbound reaction product into a new reactive structure. Each of these steps is characterized by an activation free energy.

For (*i*) the Boltzmann probability of forming a reactive state is given by $\exp(-E_R/RT)/Z$, where E_R denotes the energy of the reactive state and Z is the partition function (see Equation [1], main text). Thus, the corresponding activation energy is the difference between the free energy of the reactive state and the ensemble free energy ($-RT\ln(Z)$). This activation energy is optimized through cost function κ_1 , energies to form reactive structures are therefore lower for all PBD molecules than for CRZ-2.

For (*ii*) and (*iii*) we approximate the best refolding path from the product conformation (reactive RNA dimer) to the next reactive species in the cascade. Finding the best refolding pathways is a computationally hard problem. The best direct refolding paths (i.e. paths of minimal length) can be estimated using the *findpath* heuristic (2). In order to get a better estimate of the energy barriers, we consider not only direct refolding paths but also detours via low lying minima in the energy landscape. We computed low lying minima of RNA landscapes with the program *barriers* (3) and selected the minimum free energy (MFE) conformation and up to three of the main alternative conformations. We then computed the direct refolding paths from the product conformation to each of these low lying minima, from each low lying minimum the other and finally from each low lying minimum to the reactive structure. The barriers along direct paths are computed as the difference between the worst energy along the refolding path the energy of the starting structure, the activation barriers (*ii*) and (*iii*) are selected such that the barrier of the total path is minimal. The dissociation barrier (*ii*) corresponds to the energy needed to dissociate the cleaved end, the refolding barrier (*iii*) describes the pathway from the unbound reaction product to the new reactive structure. The resulting values show that designed molecules often have to overcome higher dissociation barriers than CRZ-2.

Table S1:Klenow primer sequences for generation of double-stranded DNA templates to be used for enzymatic synthesis of RNAs and of an inactive dimer

Test sequence Klenow primer 1 including T7 RNA promoter sequence in italic	Klenow primer 2
PBD1 5'-TAA TAC GAC TCA CTA TA GGG AGA GCA CAG TCG GAG TTG CCG CGT TAG CGG CGG TTC TAG AAG TGC CCC GCA-3'	5'-GGT TGG CAC TGA GCT TTT TCC CGC GTA ATA TAC GCC ATA TGG CTG TTT CTG CGG GGC ACT TCT AGA ACC G-3'
PBD2 5'-TAA TAC GAC TCA CTA TA GGG AGA GAA CAG TCG GTG GTG CCC CGT AAG GGG CGT CGC CAG AAG TTC GGA CCA G-3'	5'-TCG CCG ACT GTT CTT TTT GGA CCG TAA TAT ACG CCT TTT GGC TGT TTC TGG TCC GAA CTT CTG GCG ACG-3'
PBD3 5'-TAA TAC GAC TCA CTA TA GGG AGA CAG TCC GGT TTA CCG CTA ATG CGG TGG GTC GAG AAG TCT GAG CGA GAA A-3'	5'-TTT TGG TGC CGG ACT GCC TTT ATG GAG CGG TAA TAT ACC AGT ATA CTG TGT TTC TCG CTC AGA CTT CTC GAC C-3'
PBD4 5'-TAA TAC GAC TCA CTA TA GGG AGA CAG TCC GGT TTA CCG CTA ATG CGG TGG GTC GAG AAG TCT GAG CGA GAA ACA-3'	5'-TTT TGG TGC CGG ACT GCC TTT ATG GAG CGG TAA TAT ACC AGT GTC CTG TGT TTC TCG CTC AGA CTT CTC G-3'
Inactive dimer 5'-TAA TAC GAC TCA CTA TA GGG AGA GGT GTT TCA GAC TCG AGA ACC AGA GAA TGA CAC GTA TGT GCA GGA TTA ACT GGT AAA ACT CTC ACA GCT GAA ACA CCT CTT TCG G-3'	5'-GGT CTA CGA GGA TGG TCA GGA TAA GGT CGC AAG GTT GGT GGC AGC ACG CAT TAG GAC CTT GAC TTC GCT CAC AGA CCG AAA GAG GTG TTT CAG CTG TGA GAG-3'
Full length RNA sequence of the inactive dimer 5'-GGG AGA GGU GUU UCA GAC UCG AGA ACC AGA GAA UGA CAC GUA UGU GCA GGA UUA ACU GGU AAA ACU CUC ACA GCU GAA ACA CCU CUU UCG GUC UGU GAG CGA AGU CAA GGU CCU AAU GCG UGC UGC CAC CAA CCU UGC GAC CUU AUC CUG ACC AUC CUC GUA GAC C-3'	

References

1. Gan, H.H., Fera, D., Zorn, J., Shiffeldrim, N., Tang, M., Laserson, U., Kim, N. and Schlick, T. (2004) RAG: RNA-As-Graphs database--concepts, analysis, and features. *Bioinformatics*, **20**, 1285-1291.
2. Flamm, C., Hofacker, I.L., Maurer-Stroh, S., Stadler, P.F. and Zehl, M. (2001) Design of multistable RNA molecules. *RNA*, **7**, 254-265.
3. Flamm, C., Hofacker, I.L., Stadler, P.F. and Wolfinger, M.T. (2002) Barrier trees of degenerate landscapes. *Zeitschrift Fur Physikalische Chemie-International Journal of Research in Physical Chemistry & Chemical Physics*, **216**, 155-173.

Generation of circular RNA *in vivo* and *in vitro*

Sonja Petkovic, Sabine Müller

*Ernst Moritz Arndt Universität Greifswald, Institut für Biochemie, Felix-Hausdorff-Str. 4,
17487 Greifswald, Germany*

Corresponding author:

Prof. Dr. Sabine Müller
Ernst Moritz Arndt Universität Greifswald
Institut für Biochemie
Felix-Hausdorff-Str. 4
17487 Greifswald
Phone: +49 3834 8622843
FAX: +49 3834 864471
Email: smueller@uni-greifswald.de

Abstract

In the plenitude of naturally occurring RNAs, circular RNA (circRNA) and their biological role was underestimated for years, although circRNA is ubiquitous in all kingdoms of life. Circular RNA serves as genome-like material in certain viroids and in the hepatitis delta virus, occurs as (by)-product of splicing, is involved in maturation of tRNAs, and circRNA was shown being involved in regulation of gene expression by interacting with miRNAs. In addition, due to its resistance against exonucleases, circRNA is a promising tool for medicinal application, as mirrored in its ongoing application in antisense and antiviral strategies. But how is cyclic RNA formed *in vivo*, and what kinds of approaches do exist to form RNA rings with a typical 3',5'-phosphodiester backbone *in vitro*? In this review we illustrate the occurrence and mechanisms of RNA circularization *in vivo*, survey methods used for the generation of circRNA *in vitro*, and provide appropriate protocols.

Introduction

Although circular RNAs(circRNA) are found in all kingdoms of life, eukaryotes(1-7), archaea(8-12) and prokaryotes(13), their existence has been overseen for a long time. The handful of circRNAs that were known, typically were taken as a result of mis-splicing events hence representing splicing errors. However, recently the emergence of circular RNAs and their possible cellular functionalities have stronger drawn the interest of researchers, and a number of publications have shown that biological processing of RNA into circular species with still unknown function is widely spread in nature. Based on a number of recent results, circRNAs were suggested acting as post-transcriptional regulators. For example, circRNA may directly participate as miRNA antagonist in the regulation of the human tumor suppressor locus INK4/ARF(14). Similarly, the human cerebellar degeneration-related protein 1 transcript seems to be regulated by an antisense circRNA harboring 63 conserved binding sites for miRNA miR-7(15). Another circRNA containing more than 70 selectively conserved miR-7 target sites was identified in human and mouse brain(16), suggesting that these circRNAs act as highly efficient RNA-sponges. This suggestion was further supported by showing that the circular transcript of the sex-determining region Y (*Sry*), which is highly expressed in testis(2), serves as miR-138 sponge(16). Regarding these novel functionalities, it is important to differentiate between intronic and exogenic circRNAs. Intergenic or intronic circRNAs frequently occur in nature, but show only weak sequence conservation. On the opposite, exonic circRNAs were shown being abundant, stable and non-random with conserved sequences, and therefore were suggested acting as competing endogenous RNAs to bind miRNAs and/or RNA binding proteins, and thus to modulate their local free concentration(15). This suggestion gains further support by the demonstration that circRNAs can be degraded by siRNAs(17). Overall, thousands of circRNAs with often tissue/developmental stage specific expression were identified in humans, mouse and nematodes(15,17), and many more are expected to be found(18).

What makes circRNAs so attractive for cellular function? The most obvious reason is certainly stability. Since ribonucleases are ubiquitously spread, circRNAs have great advantage in terms of being protected against degradation by exonucleases *in vivo* and *in vitro*(19-22). Therefore, in addition to the still rather unknown cellular function, the covalently closed ring structure of circRNA is beneficial to a number of applications, for example as (i) antisense-RNAs, aptamers and ribozymes, (ii) templates for RNA polymerases (23), or (iii) artificial substrates for translation *in vivo* and *in vitro*(24,25). Furthermore, circRNA is a useful tool for (iv) analysis of RNA-protein interactions(20,26), as well as for medicinal purposes, for example in (v) antisense(27) and (vi) antiviral strategies(20). Thus, artificial circular RNAs have been used to induce cellular RNA interference(19,28,29).

Whether *in vivo* or *in vitro*, the question remains, what are the strategies for production of circRNA? Formation of phosphodiester linkages (ligation) in RNA proceeds by nucleophilic attack of a 3'-hydroxyl group of a first nucleotide on the (activated) phosphorus atom of a second. When circular RNA is to be formed, the 3'- and 5'-ends of the same strand have to be placed in close proximity, and impedimental secondary structures must be prevented in order to enable ligation. *In vivo*, this is mostly achieved by the RNA substrate itself that adopts a suitable conformation for intramolecular ligation, either by its intrinsic folding capacity alone, or with the help of proteins and/or other helper RNAs. *In vitro*, several

methods are available to promote the required proximity of the 3'- and 5'-terminus, and as a consequence, circularization by ligation: (i) a rod-like structure of the ligation substrate stabilized by full or partial base pairing, (ii) an oligonucleotide splint which stabilizes the region next to the ligation junction (iii) a helper oligonucleotide, which hybridizes in a distance from the ligation site, eliminating interfering secondary structures, (iv) helper proteins or aggregates of proteins and RNA, joining RNA ends, and (v) the intrinsic ligation activity of some ribozymes.

In this review we will focus on the occurrence and formation of circular RNA *in vivo* and their production *in vitro*, taking into account that 3',5'- as well as 2',5'-linkages may be formed. Furthermore, we provide a collection of protocols to be used for the purpose of RNA circularization *in vitro* (Supplementary Data).

Circularization *in vivo*

In vivo, circular RNA is often produced during the several maturation steps of pre-mRNA through splicing processes. Splicing may lead to four different products: (i) combined protein-coding RNA-regions (exons) (ii) linear non-coding RNA (intron)(iii) circular non-coding RNA with a RNA-tail (lariat) (30,31) and (iv) regular RNA circles (32). Moreover, so called mis-splicing, failure in debranching, and exon skipping may directly lead to cyclic RNAs (3,33,34). Spliceosome mediated mRNA splicing is a highly complex procedure involving more than 70 proteins, and circular RNAs are released upon exon ligation as spliceosomal introns. In similar but less complex scenarios group-I, II and III introns are formed and eventually circularized (35-38). Furthermore, autocatalytic RNA structures capable of mediating RNA circularization occur in satellite RNAs (39), in some viroids(40), and in the hepatitis delta virus (41), although in these RNA species, circularization presumably is assisted by host proteins.

a) Spliceosome

Spliceosome-mediated splicing occurs in eukaryotic protein-coding genes. The process requires small nuclear ribonucleoproteins (snRNPs)(42-45) and ATP as energy source for the organization of the spliceosome(46-48). The snRNPs provide several RNAs and more than 70 proteins that assemble to a well ordered complex structure. To be a substrate for spliceosomal splicing, a pre-mRNA with a region of non-coding sequences (intron) is required to have conserved 3'- and 5'-splice sites (49) and accompanying sequences, such as the polypyrimidine tract(50).The 3'-splice site in many cases terminates is characterized by a guanosine terminating the intron, and by the branchpoint sequence with the branchpoint adenosine located further downstream(51,52). This adenosine and its periphery are the key elements for creation of a mammalian splice intermediate, the lariat structure (49,53). Note that there are also linear intermediates such as the 5'-exon. The 2'-OH-group of the branchpoint adenosine is organized in close proximity to the 5'-exon (exon 1) utilizing the snRNPs U2, U5 and U6. U5 aligns the two exons (54), whereas U2 binds the branchpoint by base pairing, and U6 interacts with the 5'-splice site (55) (Fig. 1). Prp8 is the only protein that also interacts with the branchpoint and both splice sites, but is skipped in Figure 1 to keep it simple and clear (see (56) for a detailed review on Prp8).

The splice mechanism involves two transesterification steps as illustrated in Figure 1. The first transesterification occurs by nucleophilic attack of the 2'-OH-group of the branching nucleotide on the phosphorous atom of the upstream 5'-exon-intron phosphodiester linkage. This results in RNA cleavage followed by attack of the newly generated 3'-OH group of the 5'-exon onto the phosphorus atom of the phosphodiester that links the 3'-exon with the intron. Both transesterifications proceed as S_N2-like reactions involving an in-line attack of the nucleophile on the phosphorous(57,58). The products of spliceosome-mediated mRNA maturation are an excised lariat intron, hence a circular RNA containing a 2',5'-phosphodiester bridge, and a linear RNA composed of the two combined exons.

Fig. 1

Spliceosomal action normally leads to combined linear exons. However, *in vitro* exon circularization using mammalian nuclear extracts was also achieved. Pasman *et al.* demonstrated that splice site pairing across the exon can result in exon circularization (59), and thus provided some indication that previously hypothesized spliceosome-mediated exon circularization, accounting for rare RNAs containing scrambled exons (6,60), indeed may occur *in vivo*. Cyclization essentially depended on exon size; the observed decreased efficiency with increasing RNA chain length was overcome by using intronic complementary sequences that bring splice sites in close proximity(59). The spliceosome-mediated formation of exonic circular RNAs was further confirmed by the demonstration that a 500 kb lariat containing skipped exons upon splicing forms circRNAs independent of the number of exons or RNA size (61). Furthermore, cyclic RNAs were demonstrated being substrates for the eukaryotic translational apparatus(24,62). Schindewolf *et al.* found that an artificial circular RNA induces spliceosome assembly and splice reaction when incubated in whole cell splicing extract of *Saccharomyces cerevisiae*(62). This may also result in exon circularization and was found being independent of specific secondary structures of the circular pre-mRNA transcript. Altogether, a number of studies suggest that circular exons found *in vivo* may be generated from pre-mRNAs. In 2013, participation of exonic circRNA in gene regulation was demonstrated by identifying circular RNAs that function as microRNA sponges (63). These kinds of circRNAs are specifically and highly expressed in human embryonic kidney cells, and in human and mouse brain as well as in mouse testis(16). Already in 1995, circRNA was identified in mouse testis, and inverted repeats were found to be necessary for circularization(64). Recently Jeck *et al.* identified cyclic RNAs derived from 14.4 % of genes in human fibroblasts associated with ALU repeats (17) and proposed two models of alternative splicing for biogenesis of these cyclic RNAs (Fig. 2).

Fig. 2

b) Group-II- and group-III-introns

In addition to the spliceosomal process, autocatalytic maturation of pre-mRNA also occurs solely RNA-dependent. Despite minimal sequence and structure similarities, group-II-introns perform RNA circularization comparable to spliceosome mediated splicing(37). Instead of assistance by sRNPs for folding and exon positioning to carry out the splicereaction, group-II-intron sub-domains interact with each other to form the active complex. Only then, specific proteins are recruited for structural stabilization (reviewed in (65)). As with

spliceosomal mediated splicing, the products are a lariat involving a 2',5'-phosphodiester linkage at the branch site, and the ligated exons.

Group-III-introns splice pre-mRNA of chloroplasts in euglenoid protists. Comparable to nuclear introns, group-III-intron splicing occurs via two transesterification steps with an adenosine as initiating nucleophile. Excision of the intron results in a lariat and was discussed being evolutionary related to group-II-introns (38,66,67). For both, group-II- and group-III-introns, so far only intronic circRNA was identified.

c) Group-I-introns

In 1982, Cech and co-workers discovered that group-I-introns have an intrinsic splice activity, catalyzing their own excision from several RNA precursors(68). As with the spliceosome and group-II-introns, group-I-Intron splicing occurs via two successive transesterification steps; however, here an external nucleophile is recruited instead of the branchpoint adenosine for the first transesterification. An exogenous guanosine (G) docks onto the active G-binding site within the intron. Its 3'-OH group attacks the phosphodiester bond at the 5'-splice site, generating a free 3'-OH group at the upstream 5'-exon, whereas the exogenous guanosine becomes attached to the 5'-end of the intron. Thereafter, the 3'-terminal guanosine of the intron substitutes the exogenous guanosine and occupies the G-binding site, initiating the second transesterification. As a result, the 3'-OH group of the 5'-exon is aligned to attack the 3'-splice site, leading to ligation of the adjacent upstream and downstream exons and release of the catalytic intron (reviewed in (69)).

Already in the 1980s Cech *et al.* discovered that the excised intervening sequence (IVS) of the *Tetrahymena* RNA precursor cyclizes in an autocatalytic manner (35,36,70,71). Cyclization occurs analogous to the splice mechanism of the precursor RNA, involving cleavage of the linear IVS, 15 nucleotides from its 5'-end, by nucleophilic attack of the IVS-3'-hydroxyl terminus. As a result, the 5'-terminal sequence comprising 15 nucleotides is released, and the remaining part of the IVS is cyclized(70) (Fig. 3: a-c). Alternatively, cleavage of the linear IVS can occur 19 nucleotides from the 5'-end, leading to a minor cyclic RNA product of smaller ring size. Formation of the larger circle is more prominent and was found being pH-independent over a wide range (pH 4.7-9.0). The amount of the smaller ring was increased with increasing pH; however, yields are low compared to those of the major large circle(72). In the presence of intrinsic reactive di- or trinucleotide sequences such as UC or UCU, circle reopening may occur (see ref. (73) for a detailed *in vitro* kinetic analysis). Thus, the minor cyclization product may derive from the major larger circles by reopening, followed by nucleophilic attack of the 3'-OH terminus onto the minor cyclization site, and release of the remaining four 5'-terminal nucleotides (74). Group I intron derived circRNAs were found in a number of more organisms such as the myxomycete *Didymium iridis*(75) or *Anabaena* sp. PCC 7120(76). All of these GTP-dependent spliced linear introns may undergo further transesterifications leading to different circRNAs as depicted in Figure 3 (a-d)(77,78). Interestingly, it has been hypothesized that an evolutionary link exists between group-I-intron RNAs and satellite or viroid RNAs that also produce RNA circles, although by rolling circle replication, and show sequence similarities, *i.e.* to *D. iridis*, such as a 16 nucleotides long consensus sequence (75,79,80).

Fig. 3

Hydrolysis of the 3'-end of introns (Fig. 3: e, f) as observed for example in *Tetrahymena* is another interesting phenomenon, and it was suggested that circularization via this mechanism is a general feature of all nuclear group-I-introns (81). When circularization occurs via the 3'-end hydrolysis path (Fig. 3: e, f), exons are not combined and therefore lack functionality. Thus, the question arises: What is the benefit that had driven evolution towards circRNA, if RNA circularization occurs at the expense of functional exons? Group-I-intron derived full length circles containing the full genetic information (Fig. 3: d, e, f) were suggested playing a role in intron mobility directly at RNA level, in lateral transfer through reverse splicing. On the contrary, truncated circles that are derived from ligation of ωG to an internal residue of the linear IVS as described above (Fig. 3: a-c), might be important to prevent fatal reverse splicing. Alternatively, these circRNAs might play a role in regulation of intron homing at DNA level, by controlling expression of homing endonuclease genes (81). Often, group-I-intron derived full-length and truncated circles coexist, underscoring their functional role in the cellular life cycle.

d) Transfer and ribosomal RNA splicing in archaea

Archaeal pre-tRNA splicing involves at least two proteins: a splicing endonuclease recognizing a defined bulge-helix-bulge structural motif that is common to all archael introns, and a specific ligase. Salgia *et al.* reported on the circularization of RNA by a *Haloferax volcanii* ligase *in vivo* and *in vitro*. Endonucleolytic cleavage of the pre-tRNA produces a 5'-hydroxyl group and a 2',3'-cyclic phosphate. Thus both, exons and introns resulting from the cleavage reaction, are decorated with functionalities required for the following ligation. Accordingly, the mechanism of ligase supported exon-joining and intron circularization was found being similar. It involves a transesterification for conversion of the cyclic phosphate into linking phosphate, resulting in a (typical) 3',5'-phosphodiester bridge(11,82). Archaeal endonucleases are also involved in rRNA splicing. Thus, cleavage and exon-joining reactions for rRNAs in archaeabacterium *Desulfurococcus mobilis* resemble those found for pre-tRNA splicing. The rRNA intron circularizes by intramolecular ligation, very likely employing the same enzyme that is also used for exon ligation(10). Notably, the 23S rRNA of another archaea *Pyrobaculum organotrophum* contains two introns that after excision circularize and form a 3',5'-phosphodiester linkage (83).

e) RNA self-cleavage and ligation activities beyond RNA introns

Beyond catalytic introns, there are a number of other RNAs, as for example the hammerhead or hairpin ribozyme, which perform cleavage and/or ligation reactions by transesterifications. In general, the self-cleaving/self-ligation activity of ribozymes *in vivo* may be involved in (i) processing of replication intermediates (hammerhead, hairpin, varkud satellite, hepatitis delta virus ribozyme), (ii) control of gene expression (gIms ribozyme), (iii) maturation of t-RNA (RNase P) and (iv) maturation of pre-mRNA(see above: splicing)

The origins of self-cleaving and -ligating RNA activities are found in viroids, viruses and satellite RNAs. The first discovered satellite RNA derived self-cleaving RNA structure was extensively investigated, modified and minimized, and the resulting catalysts were named hammerhead ribozymes, due to their striking secondary structure. Another small catalytic RNA, the hairpin ribozyme, has the same origin as the hammerhead ribozyme, as it is derived from viral satellite RNAs, e.g. from tobacco ringspot virus (sTobRV). The biological function of both is to cleave multimeric products of rolling circle replication to monomeric

linear "genoms"(84,85) (Fig. 4). Since viroid, virus and satellite RNAs may occur in linear and cyclic states *in vivo*, it was also postulated that hairpin and hammerhead motifs ligate linear to circular RNA. The underlying mechanism *in vivo* is not fully understood, but isolates of sTobRV were discussed to perform cleavage and ligation via nucleophilic attack of the 5'-OH group on the terminal 2',3'-cyclic phosphate, whereby the global RNA fold favors a trigonal bipyramidal intermediate (39).

Fig. 4

Hepatitis- δ -virus (HDV) is a single-stranded circular RNA satellite virus that requires co-infection of the host with Hepatitis B virus (HBV) for its life cycle. It replicates via the double rolling circle mechanism (reviewed in (86)) (comparable to Fig. 4) and is the fastest natural self-cleaving RNA catalyst (87). The plus-strand (genome) is the template for the human host RNA polymerase I (RNAP I) creating a linear multimeric negative strand. The produced RNA strand is autocatalytically cleaved in monomeric linear strands of defined length. The monomer folds into a rod-like highly double-stranded form (88,89), bringing the 3'- and 5'-ends in close proximity. Ligation leads to the circular "antigenome". The second step of replication is accomplished by host RNA polymerase II. Again self-cleavage of linear multimers in monomers, formation of the rod-like double-stranded structure and ligation are performed, reproducing the circular genome. Different reports suggest that circularization is conducted by a host ligase(90) associated with viral antigens (91), or alternatively, by ribozymatic self-ligating activity (92).It was reported that at least subfragments of HDV-RNA are able to self-cleave(93) and self-ligate in dependence on the concentration of divalent cations as for example Mg²⁺(94). Contradictory, Ke *et al.* discussed structural features including an "active site collapse" that renders the reverse reaction of cleavage impossible(95). Although the sequence of the HDV ribozyme does not compare to the hairpin or hammerhead motif, a 5'-OH-group and a 2',3'cyclic phosphate are involved in circularization as described above(90,92).

f) Viroids

Viroids consist of single-stranded (ss) circular non-encapsidated RNAs of 250-400 nucleotides length, which are pathogenic to plants. The RNA of viroids is noncoding and nontranslatable(96). How was it possible that this kind of RNA survived in early life development and how does it survive now, in a cellular environment? Is the solution simply RNA circularity or the formation of double-stranded regions or both?

It is possible to distinguish between three different types of viroids: (i) *Avsunviroidae* possessing RNA that consists of a Y-shaped or branched circular structure(97,98) replicating in the chloroplast (99); (ii) *Popsiviroidae* forming a rod-like circular structure (100) and replicating in the nucleus and nucleolus; and (iii) the so called retroviroids, which contain branched RNA and replicate in the nucleus. All three viroidae groups replicate via the rolling circle mechanism.

The mechanism of RNA circularization in *Avsunviroidae* is not yet completely understood. There are three possible scenarios reviewed in detail by R. Flores(40,41). In short: Intramolecular ligation may be performed enzymatically by a plant ligase using the 5'-OH-group and 2',3'-cyclic phosphate generated by cleavage of the multimeric RNA strand produced from rolling circle replication. Alternatively, self-ligation by hammerhead

ribozyme structures was discussed. Very interestingly, peach latent mosaic viroid transcripts form rod-like structures juxtaposing ends that self-ligate (101) producing >96% of the 2',5'-isomer (102). The 2',5'-phosphodiester bond was hypothesized enhancing the stability of the formed circular RNA by preventing further viroid self-cleavage, whereas it does not prevent further replication of the viroid (102). *Popsiviroidea* possess a rod-like RNA structure, optimal for circularization by RNA ligase. Thus, these viroids seem to process the RNA to circles utilizing a host RNA ligase (103), also reviewed in (40,41). Altogether, viroids as circular RNAs with large double stranded regions seem to profit from both, circularity and double stranded structure. They show resistance against nucleolytic decay and may only be hydrolyzed by endoribonucleases; wherein only endoribonucleases with dsRNA specificity may play a role. However, dsRNA specific ribonucleases are not widely spread in nature. Thus, circularity and eventually additional formation of double stranded structures is very beneficial in particular in terms of enhanced stability, and therefore may have led to RNA species survival in early life (104) as well as in cellular environment.

Circularization *in vitro*

RNA circularization *in vivo* leads to 2',5'- or 3',5'-phosphodiester linkages. There are mechanisms requiring protein assisted (e.g. spliceosomal circularization) or protein-independent (e.g. group-II-introns or small ribozymes) RNA activity. Although natural circRNAs might consist of double stranded regions, the initial product of ligation is steadily a single stranded circRNA. Except for (pre-) tRNAs, little RNA modifications were observed in naturally occurring RNA circularization processes. Therefore it is unknown whether splicing, ribozymatic or other *in vivo* circularization strategies would tolerate RNA nucleobase or backbone modifications. Synthetic methods instead mainly depend on ligation site chemistry and may therefore be used when modified circRNA is to be formed.

Chemical methods

There are two major agents that have been used for ester bond formation between 3'-hydroxyl and 5'-phosphate groups of nucleic acids: cyanogen bromide (BrCN) and ethyl-3-(3'-dimethylaminopropyl) carbodiimide (EDC)(105-107). Since cyanogen bromide possesses higher reactivity (105) and was more widely used, we here discuss only BrCN assisted ligation in RNA circularization experiments.

Damha and Carriero successfully circularized branched RNAs to lariats with the help of BrCN and 2-(N-morpholino)-ethane sulfonic acid (MES), to prepare substrates for studying the mechanism of a debranching enzyme(108). To obtain exclusively 3',5'-phosphodiester linkages, deoxythymidine was used as 3'-terminal nucleotide. The authors made use of bimolecular association, generating double stranded regions with a DNA or RNA splint (helper strand) juxtaposing the ends to be ligated. Yields in the range of 45% were obtained (108), using the protocol provided in Supplementary Data, Protocol S1.

Wang and Kool reported on the synthesis of non-lariat circular RNAs(106,109,110). Typically, chemical ligation of full RNA substrates leads to mixtures of RNAs consisting of isomeric mixtures of 2',5'- and 3',5'-phosphodiester bonds. Kanaya and Yanagawa found that addition of Ni²⁺ shows the highest ratio of 3',5'-phosphodiester to 2',5'-phosphodiester bond

formation compared with Co^{2+} , Zn^{2+} and Mn^{2+} (106), and thus Ni^{2+} was used preferentially. Alternatively, as mentioned in the example above (108), a deoxyribonucleotide at the 3'-end was used to guarantee uniformity of 3',5'-phosphodiester bond formation at the ligation site(107). The termini were aligned using a deoxyoligoribonucleotide splint in low excess over the pre-circRNA. The yield of circularization was 75-85% (protocol provided in Protocols2, Supplementary Data). Mechanistically, BrCN activates the 3'-phosphate to be attacked by the 5'-hydroxyl-group acting as nucleophile (Fig. 5).

Fig. 5

Fedorova et al. recommended a protocol for chemical ligation of DNA, using morpholino derivatives of higher activity compared with MES, thus obtaining 80% of a circular oligonucleotide(111). This procedure might be transferred to RNA circularization protocols. In 1999, Micura and coworkers reported on a novel approach for solid phase synthesis of cyclic RNAs of 2-21 nucleotides length (112). The procedure combines standard phosphoramidite chemistry for chain assembly and phosphotriester chemistry for ring closure. After chain assembly was finished, final detritylation generated a 5'-terminal hydroxyl group for nucleophilic attack on the unique 3'-terminal phosphodiester, leading to circularization while the oligonucleotide was still linked to the polymer. Here, instead of BrCN or carbodiimide, 1-mesitylenesulfonyl-3-nitro-1,2,4-triazole was used as condensing agent. Subsequent cleavage of cyanoethyl groups enabled selective detaching of cyclic RNAs, because now only the circRNAs are bound through phosphotriesters via an *o*-chlorophenyl group to the resin and therefore could be selectively removed using tetramethylguanidinium pyridine-2-aldoximate. Remaining non-cyclized or mis-linked oligoribonucleotides were not removed due to their phosphodiester linkages. Circularization was followed by removal of the remaining base- and phosphateprotecting groups and RNA purification. This method led to cyclized RNA with an average yield of 15% (112).

Enzymatic methods

T4 RNA ligase 1

T4 RNA ligase 1, isolated from bacteriophage T4, catalyses circularization of RNA using a 5'-phosphoryl-terminated nucleic acid donor and a 3'-hydroxyl-terminated nucleic acid acceptor in an ATP-dependent reaction. Generally, all RNA strands in a reaction mixture provide these ends, such that intermolecular ligation is likely (113-115), but the predominant event is intramolecular reaction and thus formation of a ring from a single-stranded RNA (116,117). The RNA chain must be at least eight nucleotides long for circularization (114) using the protocol provided in Protocols3 (Supplementary Data).

Romaniuk et al. examined the best composition of a trinucleotide bisphosphate acceptor by testing each individual position. In their experiments, the authors did not focus on circularization but ligation in general. Comparing pCp and pUpUpUpCp as donors, it was found, that there are preferences for the nucleotide providing the reactive 3'-hydroxyl group in the following order: A > G ≈ C > U. The second position follows the tendency: A > C > U > G, implying that highest ligation yields are obtained using two adenines at the 3'-terminus of the acceptor. For the donor, the situation is less clear, but pyrimidines are

preferred over purines (118). Whereas the donor must be activated preferably through adenylation, the 3'-OH group is mandatory for successful ligation(119).

The length of the acceptor is not as important as the base composition since a series of oligo (U)s decreases and a series of oligo(A)s increases the ligation yield dramatically. Ligation yield can be accelerated, by reduced MgCl₂ concentrations, or by adding dimethyl sulfoxide as a stabilizing agent (120). Additions of other stabilizing agents such as urea, Carbowax or formamide, or the use of sodium or potassium salts did not enhance reaction yield at a fixed enzyme concentration.

The mechanism starts with activation of the enzyme via adenylation using ATP (Fig. 6: (A)). An adenylate transfer to the donor sequence takes place, forming an A(5')pp(5')D intermediate, and the enzyme is released (Fig. 6: (B)). Finally, the free enzyme catalyses ligation, linking the intermediate with the 3'-OH group of the acceptor sequence and releasing AMP. Assayed features of T4 RNA ligase 1 such as, Mg²⁺-dependency, stoichiometry of the reaction, substrate nature, pH-range, chain length and reaction inhibitors can be looked up in Ref. (117). Exemplary protocols for RNA circularization with T4 RNA ligase 1 are shown in Supplementary Data (Protocols S4(114) and S5(121)).

Fig. 6

T4 RNA ligase has been used for circularization in a number of applications. For example, experiments utilizing T4 RNA ligase 1 yielded up to 80% circular hammerhead ribozymes at 30 μM RNA concentrations (120). However, this method did not work for all substrates used in this study; some were preferentially ligated to linear dimers. A decreased RNA concentration from 30 to less than 1 μM was described as being favorable for ring formation; however, obtained yields were not higher than 5%. To increase circularization yields, Wang and Ruffner used a helper deoxyoligonucleotide, which hybridizes in a distance from the reactive ends. Depending on the ligation substrate, helper strands folded as hairpins were used as well as linear DNA forming a hybrid double-strand with the substrate (Fig. 7a).

Fig. 7

Both types of helper deoxyoligonucleotides assisted folding of the RNA substrate and pre-orientation of the reactive ends, and as a consequence, improved the yield of circularization for distinct substrates. Use of the linear helper oligonucleotide increased the yield of circRNA to 40%, whereas the hairpin helper allowed for ring formation with nearly 100% yield(122). Not surprisingly, increased RNA concentrations decreased the yield of the cyclic product, even when the helper strand was added. Interestingly, all cyclic ribozymes synthesized in this study displayed increased catalytic activity, resistance against exonucleolytic degradation, and decreased need of divalent cations(122)(Protocol S6, Supplementary Data). Helper hairpins turned out being particularly beneficial to the ligation of complex RNA structures. They promote ligation because tangled units may be eased, driven by the free enthalpy obtained through each DNA-RNA base pair. It is likely that polynucleotides may be ligated using this kind of helper since the increasing number of base pairs favours the enthalpic term of the reaction. However, if the strands are too long, the positive effect will be minimized when additional secondary structures in the helper sequence occur. The loop might become too long and flexible, favouring unwanted

secondary structures. Therefore, if unexpected Watson-Crick, Hogsteen and sugar edge interactions are involved, instead of standard Watson-Crick base pairing, there is no advantage in using the “helper” although lots of enthalpy will be released.

Instead of an external DNA helper oligonucleotide strand, internal base pairing may promote ligation(123). Accordingly, Abe *et al.* designed dumbbell shaped nanocircular RNAs for RNA interference purposes (fig. 7b) and demonstrated successful circularization by T4 RNA ligase1, closing the loops of substrates in the range of 15 to 27 base pairs (19). In the following years, the authors reported several more successful RNA circularizations using T4 RNA ligase 1 (28,124-126). An exemplary ring formation protocol following this strategy is shown in Protocol S7 (Supplementary Data).

Sugino *et al.* found that the smallest intermolecular reaction utilizing bacteriophage T4 RNA ligase requires 2 to 3nt, and the minimal sequence length for a circularization is 6nt(119). However, the optimal length for ring formation is 20nt. Longer sequences may include involuntary structures, such that the design and application of a helper strand like the one suggested by Wang and Ruffner as explained above(122) may be additionally complicated. For large RNAs in the kb range, other methods like the permuted exon-intron method PIE (127)(see below) also termed "RNA cyclase ribozyme"(128) may be advantageous.

For generation of the dumpbell shaped nanocircles, a 5'-phosphorylated RNA was circularized. 5'-phosphorylation is achieved enzymatically using a kinase, or by chemical synthesis. What about enzymatically derived RNA substrates? They typically carry a triphosphate at the 5'-end, that does not allow for circularization directly. One possibility is the hydrolysis of the triphosphate followed by a kinase reaction (129)(130)(Supplementary Data, Protocol S8). To enhance ligation, joining ends were juxtaposed due to a Watson-Crickbase pairing stem (Fig. 7c top), and branched 80 nt long single-stranded RNA was ligated using T4 RNA ligase 1 following the protocol provided in Protocol S9 (Supplementary Data)(130). Alternatively, a deoxyribozyme was used for circularization (Fig. 7c bottom)(131), but the afore mentioned two-step approach turned out being more efficient delivering the circRNA in higher yields.

To circumvent the two enzymatic steps of de- and rephosphorylation, GMP-primed *in vitro* transcription is an elegant way for obtaining 5'-monophosphorylated linear RNA substrates (132,133). This strategy was successfully used for RNA circularization by T4 RNA ligase 2(134) or T4 DNA ligase (24).

T4 RNA ligase 2

Ho and Shuman found bacteriophage T4 RNA ligase 2 (in short Rnl2), which possesses also inter- and intra-molecular ligation activities (135). Also for this ligase, intramolecular ligation of ssRNA is the predominant reaction. The mechanism is comparable to the mechanism of T4 RNA ligase 1 (see above), passing three steps and forming adenylated enzyme and substrate intermediates (for detailed structural analysis of RNA ligase 2 and comparison to DNA ligase 1 see (136,137)).

Rnl2 has also significant activity for ligation of double-stranded (ds) RNA. This feature can be used to ligate RNA with a nick in a double stranded region to form a cyclic RNA(134). T4 RNA ligase 2 also accepts DNA as the second strand when a hybrid composed of DNA and a nicked RNA is to be ligated. Phosphodiester linkage formation was studied using different divalent cations (Mg^{2+} , Mn^{2+} , Zn^{2+} , Ca^{2+} , Cu^{2+}) and nucleoside triphosphates (ATP, GTP, UTP, CTP and dATP)(138)(see Protocol S10 for a protocol of ligation of a nicked double stranded

region). It turned out, that magnesium and manganese ions are suitable for nick ligation and cannot be substituted by zinc, calcium or copper ions. Also ATP usage is indispensable. A detailed mechanistic and kinetic analysis of T4 ligase 2 can be found in Ref. (135). There is a homologous protein to T4 RNA ligase 2 occurring in vibriophage KVP40. It also catalyzes RNA circularization of single stranded small (such as 18mers) RNAs in an ATP-dependent manner, if RNA is not adenylated (see(139) for detailed comparison and kinetic data of T4 and KVP40 RNA ligase 2 dependent circularization). Furthermore, a truncated version of T4 Rnl2 with 249 amino acid residues is commercially available. It ligates the preadenylated 5'-end of DNA or RNA to the 3'-end of RNA. No further ATP is required (135,136).

tRNA ligases from wheat germ, Arabidopsis and S. cerevisiae

To our knowledge the following (t)RNA ligases from wheat germ, *Arabidopsis* and *S. cerevisiae* are not commercially available but should be at least mentioned here since these ligases may form circular RNA, too(140,141). The mechanism of action of T4 RNA ligase (described above) and wheat germ ligase is quite similar but the substrate requirements differ. Abelson *et al.* demonstrated that wheat germ ligase preparations are able to ligate RNA with a 5'-phosphorylated or a 5'-hydroxyl terminus. The preparation also consisted of a 5'-hydroxyl polynucleotide kinase which might generate the 5'-phosphorylated terminus. However, also 2',3'-cyclic phosphates or 2'-phosphates were found being accepted as substrates by wheat germ ligase(140,142). In the plant, this ligase catalyses the splicing of pre-tRNA and generates a circular RNA as by-product (143). A similar ligase was found in yeast (141,144). A recombinant yeast RNA ligase was shown to convert linear introns to their cyclic counterparts, although notably slower than wild-type *Arabidopsis* tRNA ligase (145).

Thermostable RNA ligase from Pyrococcus abyssi

Another RNA ligase was found in *Pyrococcus abyssi*, termed Pap1020. It performs the ATP dependent (no other nucleotide) circularization of exclusively oligoribonucleotides (no deoxyribonucleotides). Ring formation occurs by intramolecular reaction between the 3'- and 5'-termini of the RNA, whereby in contrast to T4 Rnl1 or 2 no oligomerization was observed(146). A wide range of temperatures from 20 to 95 °C were tested and found suitable for efficient ligation. Testing of several ATP concentrations ranging from 5 µM to 1 mM revealed that circularization activity decreased with increasing ATP concentration (see Supplementary data, Protocol S11 and Ref.(146)).

GTP-dependent Escherichia coli RNA ligase

A specific ligase was described by Tanaka, named RtcB(147). It belongs to an *E. coli* RNA repair operon with still unknown mechanistic features. In 2012, Chakravaty *et al.* described the three step mechanism of RNA ligation and circularization(148): Intramolecular ligation is conducted when the ligase first binds GTP via a histidinyl residue, creating a histidinyl-N-GMP intermediate. A polynucleotide-3'-phosphate takes over the guanylate and forms a polynucleotide-(3')-pp-(5')G intermediate. The 3',5'-phosphodiester bond is formed by nucleophilic attack of the 5'-hydroxyl group onto the activated 3'-phosphate of the intermediate. This mechanism differs from others by activating a 3'-monophosphate as

phosphoanhydride with GMP. The authors tested several GTP concentrations for RNA ligation and circularization, finding increasing yields of circRNA with increasing GTP concentration, up to complete ligation at 6.25 µM GTP (see Protocol S12, Supplementary data).

T4 DNA ligase

T4 DNA ligase, also derived from Bacteriophage T4, catalyzes the formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini in duplex DNA or RNA. The enzyme joins blunt and cohesive termini as well as repairs single-stranded nicks in duplex DNA, RNA or DNA/RNA hybrids(149).

Ligation by T4 DNA ligase leads to circular RNA as does ligation with T4 RNA ligases using a similar mechanism(150-152), but the substrate requirements differ. Generally, T4 DNA ligase is used when dsRNA is to be ligated or circularized. However, ssRNA rings can be formed, when a helper deoxyoligonucleotide (splint) is used. Ligation should be performed using RNA and splint in equimolar amounts, and if the RNA is highly structured, longer splints up to full length base pairing may be advantageous, in combination with lower temperatures. The ends of the RNA which is to be circularized or ligated and the splint form a ligation competent complex (LCC) via base pairing, before T4 DNA ligase performs its reaction (153) (Fig. 8a).

Fig. 8

An interesting example of splinted RNA circularization was published already in 1995. Chen and Sarnow reported the circularization of a 453nt long RNA with T4 DNA ligase, to be used for the analysis of initiation of protein synthesis by the eukaryotic translational apparatus on circular RNA transcripts(24)(Supplementary Data, Protocol S13). It was shown that the *in vitro* derived circRNA allows for eukaryotic protein translation if the internal ribosome entry site is included in the circRNA's sequence.

In addition to DNA, also RNA splints can be used. Lang and Micura employed a 2'-O-methylated RNA splint to ligate a single stranded RNA. The nick was positioned in a loop (Fig. 8b), and it was closed using T4 DNA ligase(154). 2'-O-methylated RNA enhances the stability of the splint-substrate complex, and thus may help to improve ligation. Note that the splint should not be phosphorylated to reduce unwanted splint circularization or multimerization. In the majority of publications, protocols for T4 DNA ligase as well as RNA ligase mediated circularization of short RNAs (< 500nt) are described. It is evident that temperature, stoichiometry and splint length have to be adjusted for each substrate. Large substrates in kb scales require different strategies, due to structural challenges. If the splint is longer, the likelihood is higher that also in the helper molecule unfavourable structures will occur. Therefore, as mentioned above, for circularization of large RNAs protocols using the PEI strategy (described below) may be the better alternative.

Artificial rolling circle replication

As explained above, the rolling circle mechanism (rcm) is a unidirectional process of nucleotide replication and proceeds via three steps: (i) synthesis of a multimeric transcript from a cyclic template by a RNA polymerase; (ii) cleavage to unit-length, autocatalytically or using *trans*-active ribozymes; and (iii) circularization by a ligase or autocatalytic

activity(155). Also described above, viroids, virusoids(156) and hepatitis delta virus (86) use the rcm to replicate their RNA. The templates can be single or double-stranded RNA.

An application using the rolling circle mechanism and the self-splicing activity of the hairpin ribozyme for the generation of circular RNA was described in 1998. Kool and Diegelman utilized circular single stranded DNA templates for *in vitro* transcription by either T7 or *E.coli* RNA polymerase(157). *In vitro* transcription results in multimeric repeating RNAs including the hairpin ribozyme element. Therefore, upon transcription RNA cleaved itself in monomer-length segments. The authors observed the following formation of circular monomers and suggested that circularization occurs because of the intrinsic hairpin ribozyme activity (158).

Data from our laboratory and from Dallas *et al.* supports the idea of hairpin ribozyme derived RNA circularization (134)(159). We have demonstrated that specifically designed linear RNAs harboring the hairpin ribozyme structure undergo self-processing and circularization (Fig. 9). In addition to cyclic species derived from linear monomers, also cyclic and linear dimers and oligomers were identified (134).

Fig. 9

In 2006, Kazakov *et al.* showed hairpin ribozyme circularization at freezing conditions(160). Working in frozen solution, e.g. at -10 °C and pH 8.0, increased ligation rates up to 65% were achieved. Under these conditions, an independency of divalent cations for reaction was found, but monovalent cations (always with chloride as counterion) varied in their influence on ligation. Circularization was supported in the following order: LiCl > NH₄Cl > NaCl >> KCl ~ Tris-Cl. Other counter ions tested with Na⁺, such as fluoride, bromide, iodide, nitrate, or perchlorate had no or even a negative effect on intramolecular ligation. The authors also investigated reaction in unfrozen solutions and the effect of ethanol addition. In the range of 40 to 60% ethanol, reaction produced circular RNAs independent of the presence or absence of Mg²⁺. Similarly, evaporation of water from diluted aqueous solution at 25 °C supported circularization. In addition, RNA species not observed under the conditions described above were formed, and it was suggested that those species might be multimers. Addition of 2-6% PEG to frozen, evaporated, or ethanolic solution increased the yield of ligation products, whereas PEG alone did not support circularization.

Permuted Intron-Exon (PIE) method (RNA cyclase ribozyme)

The methods of chemical or enzymatic circularization of linear RNAs are multi-step procedures requiring the synthesis, functionalization and purification of the linear RNA followed by synthesis and purification of the circular RNA. As a less tiring alternative, self-splicing introns (group-I or II) may be used to circularize exons *in vitro*. This is particularly attractive, because the method may be adaptable to spliceosomal introns suitable for specific cell types or tissues.

Careful analysis of the pre-rRNA of *Tetrahymena* created the basis for bringing splice sites into service. Already in 1987, it was found that a 5'-splice site may function not only upstream, but also downstream from the 3'-splice site (161). Been and Puttaraju summarized the functions of splice sites to be defined for one by the splice site sequence, and for second by structural characteristics which (i) recognize, (ii) bind, and (iii) activate the splice sites. So it is feasible to generate circles by reversing splice site sequences (127), and

the general procedure is named permuted intron-exon strategy (PIE). Originally, this strategy was introduced at the example of *Anabena* and *Tetrahymena* sequences in 1992 (127). Three years later Ford and Ares rearranged the *td* gene intron from phage T4 and termed the resulting transcript that upon two transesterifications produces circRNA “cyclase ribozyme”(128). Typically, in the natural group-I-splice mechanism, two exons are combined in the absence of protein components. For PIE, the intron is reversed, halved, and combined with a reversed assembly of exons (Fig. 10). The first transesterification leads to release of the 3'-terminal sequence (5'-half intron) of the PIE construct. The newly generated free 3'-OH group of exon 1 attacks the 3'-splice site (3'-ss) in the second transesterification. This results in circRNA and release of the 3'-halve intron (Fig. 10).

PIE can be applied *in vitro* as well as in several organisms, since this procedure generally requires only Mg²⁺ and guanosine nucleotides. In addition, Puttaraju and Been found that the PIE method is more efficient than natural splicing at higher temperatures (Table 1).

Table 1

One of the most recent examples of circRNA formation via PIE is the construction of miRNA sponges (162) following the strategy illustrated in Figure 10.

Fig. 10

The PIE method was used to synthesized circular RNA aptamers *in vitro* and *in vivo*(21), as well as to generate circular hammerhead (163) and hepatitis delta virus ribozymes (164). Production of RNA circles in *E.coli* as well as in yeast verified experimentally that construction of rearranged introns enables circularization of RNA *in vivo*. Basically, it is possible to create any desired circular RNA if the sequence is inserted into a particular site in the plasmid, creating a new RNA cyclase ribozyme gene. Circular RNA in the range of 71 to 1130 nt were generated. There are many more examples of application of PIE. Reference(165) provides a detailed summary of *in vivo* and *in vitro* circular RNA expression, several different exon and intron targets, strategies and further references.

Modified group-II-introns

It is also possible to modify group-II-introns for inverse splicing *in vitro* generating RNA circles (166,167). Mikheeva *et al.* succeeded in generating human cyclic exons without the assistance of any heterologous sequences. For doing so, they engineered a derivative of a yeast self-splicing group II intron that catalyzes the formation of a circular human exon *in vitro*. RNA circles were formed by inverse splicing, which requires arranging the exons consecutively, positioning the branchpoint upstream and intronic sequences up- and downstream of the exons (Fig. 11). This design allows exon circularization upon two transesterifications, excluding any intronic sequences in the circle. An advantage of this strategy in comparison with group I intron mediated exon circularization is the higher efficiency, the production of circRNA being composed entirely of exonic sequences, and the complete sequence variability (for group I intron mediated exon circularization the 3'-terminal residue of the 5'-exon must be U). Note, that circRNAs produced via this pathway carry a 2',5'-phosphodiester at the ligation/circularization site.

Fig. 11

Summary

Circular RNA occurs in all life-forms, eukaryotes, prokaryotes and archaea, as well as in proto-life forms like viroids, and (helper) viruses. Thus, evolution itself demonstrates the natural importance of circular RNA. Moreover, circRNA appears in many cellular processes, such as splicing, RNA replication, or RNA processing involving cleavage and ligation, and circRNA was suggested being involved in the regulation of gene expression by interacting with miRNAs as micro RNA sponges(16,63,162). Being more stable than its linear counterpart, circular RNA is a likely candidate for storage of genetic information in early life forms, today still represented in viroids and virus-like particles. Accordingly, circRNA recently was discussed as a possible cyclic chromosome in a proto-cell referring to RNA world scenarios (168). RNA circularization once might have occurred randomly by the intrinsic catalytic activity contained in some RNAs. In modern life, ligation *in vivo* is mediated by proteinaceous ligases, but also by still existing ribozyme activity.

Encouraged by the discovery of ever more cyclic RNA species *in vivo*, there has been also growing interest in strategies for RNA circularization *in vitro*. Suitable models are needed for studying the structural and functional features of cyclic RNA occurring *in vivo*. Furthermore, applications in molecular biology, diagnostics or medicine will profit from the higher stability of circRNA compared to their linear counterparts. There are a number of strategies for circularization of RNA to a covalent closed ring *in vitro*. Interestingly, apart from chemical attempts, experimenter's strategies mimic biological procedures, for instance as described for the rolling circle mechanism or the manipulation of group-I- or group-II-intron-splicing. In addition, commercially available ligase enzymes or ribozymes can be used for the generation of circRNA. Therein, handling of pure RNA experiments, hence using RNA mediated circularization strategies, may be easier than combination of several classes of biomolecules in one pot.

Dependent on the size and nature of the circular RNA to be prepared, the appropriate circularization method may be chosen. For example, modified group-I- or group-II-introns as used in the PIE protocol or inverse splicing, do not allow for insertion of modified bases and/or custom-made sugar-phosphate backbones. In a similar manner ligation by proteinaceous ligases or ribozymes is limited to natural nucleosides and even particular sequences at the ligation/cyclization site. This problem can be circumvented by chemical RNA ligation using BrCN orcarbodiimides, or by direct solid phase synthesis of circRNA as demonstrated by Micura and co-workers. As a compromise, RNA fragments carrying modifications may be chemically synthesized, and provided that modifications are not located at the ligation junction or are tolerated by the ligase, can be enzymatically ligated. In general, the choice of the method should be based on three key points: the size of the circRNA, its nature (natural or modified, ss or dsRNA), and the site of production (*in vitro* or *in vivo*). Table 2 provides an overview on established protocols and a recommendation for the circularization strategy for an individual target. Taken together, there are circularization strategies for all kinds of RNAs. Researchers still discover new ribozymes and ligases, and therefore the potential for RNA circularization is not exhausted.

Table 2

Following the research over the past 20 years, circular RNA has gained drastic interest in 2013.

Fig. 12

Artificial circRNA is getting more and more involved in therapeutic applications. Umekage and Kikuchi demonstrated several times, that production of circRNA aptamers using PIE *in vivo* and *in vitro*, is well possible(21,165). There are many aptamers with potential for therapeutic application that might profit from circularization. For example, a RNA aptamer targeting a specific strain of influenza A virus subtype H3N2 was shown having a 15-fold higher affinity to haemagglutinin compared with a corresponding monoclonal antibody (169). If RNA circularization and this antiviral strategy are going to be combined, it may be possible to generate a stable circular aptamer for *in vivo* application. In general, using circular RNA instead of monoclonal antibodies for targeting viral surface proteins may lead to a change of thinking how to treat viral infection. In addition, circRNAs were successfully used in anticancer strategies applying miRNA sponges in HeLa and SK-MEL cell lines against malignant melanoma(162). Although it is still a long way from cell line experiments to real medical application, the use of circRNA is a highly promising option, not at least since there are many methods to generate any kind of circRNA, and one may anticipate that the toolbox is still growing.

References

1. Bailleul, B. (1996) During *in vivo* maturation of eukaryotic nuclear mRNA, splicing yields excised exon circles. *Nucleic Acids Res*, **24**, 1015-1019.
2. Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovell-Badge, R. (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, **73**, 1019-1030.
3. Cocquerelle, C., Mascrez, B., Hetuin, D. and Bailleul, B. (1993) Mis-splicing yields circular RNA molecules. *FASEB J*, **7**, 155-160.
4. Daros, J.A. and Flores, R. (1996) A group I plant intron accumulates as circular RNA forms with extensive 5' deletions *in vivo*. *RNA*, **2**, 928-936.
5. Li-Pook-Than, J. and Bonen, L. (2006) Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res*, **34**, 2782-2790.
6. Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1991) Scrambled exons. *Cell*, **64**, 607-613.
7. Matsumoto, Y., Fishel, R. and Wickner, R.B. (1990) Circular single-stranded RNA replicon in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **87**, 7628-7632.
8. Burggraf, S., Larsen, N., Woese, C.R. and Stetter, K.O. (1993) An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci U S A*, **90**, 2547-2550.
9. Dalgaard, J.Z., Garrett, R.A. and Belfort, M. (1994) Purification and characterization of two forms of I-Dmol, a thermophilic site-specific endonuclease encoded by an archaeal intron. *J Biol Chem*, **269**, 28885-28892.

10. Kjems, J. and Garrett, R.A. (1988) Novel splicing mechanism for the ribosomal RNA intron in the archaeabacterium *Desulfurococcus mobilis*. *Cell*, **54**, 693-703.
11. Salgia, S.R., Singh, S.K., Gurha, P. and Gupta, R. (2003) Two reactions of *Haloferax volcanii* RNA splicing enzymes: joining of exons and circularization of introns. *RNA*, **9**, 319-330.
12. Starostina, N.G., Marshburn, S., Johnson, L.S., Eddy, S.R., Terns, R.M. and Terns, M.P. (2004) Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc Natl Acad Sci U S A*, **101**, 14097-14101.
13. Molina-Sanchez, M.D., Martinez-Abarca, F. and Toro, N. (2006) Excision of the *Sinorhizobium meliloti* group II intron Rmlnt1 as circles in vivo. *J Biol Chem*, **281**, 28737-28744.
14. Burd, C.E., Jeck, W.R., Liu, Y., Sanoff, H.K., Wang, Z. and Sharpless, N.E. (2010) Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet*, **6**, e1001233.
15. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M. et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333-338.
16. Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384-388.
17. Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. and Sharpless, N.E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, **19**, 141-157.
18. Wang, P.L., Bao, Y., Yee, M.C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinneny, J.R., Brown, P.O. and Salzman, J. (2014) Circular RNA Is Expressed across the Eukaryotic Tree of Life. *PLoS One*, **9**, e90859.
19. Abe, N., Abe, H. and Ito, Y. (2007) Dumbbell-shaped nanocircular RNAs for RNA interference. *J Am Chem Soc*, **129**, 15108-15109.
20. Bohjanen, P.R., Colvin, R.A., Puttaraju, M., Been, M.D. and Garcia-Blanco, M.A. (1996) A small circular TAR RNA decoy specifically inhibits Tat-activated HIV-1 transcription. *Nucleic Acids Res*, **24**, 3733-3738.
21. Umekage, S. and Kikuchi, Y. (2009) In vitro and in vivo production and purification of circular RNA aptamer. *J Biotechnol*, **139**, 265-272.
22. Harland, R. and Mishler, L. (1988) Stability of RNA in developing *Xenopus* embryos and identification of a destabilizing sequence in TFIIIA messenger RNA. *Development*, **102**, 837-852.
23. Sarah L. Daubendiek, K.R., Eric T. Kool. (1995) Rolling-Circle RNA Synthesis: Circular Oligonucleotides as Efficient Substrates for T7 RNA Polymerase. *Journal of American Chemical Society*, **717**, 7818-7819.
24. Chen, C.Y. and Sarnow, P. (1995) Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science*, **268**, 415-417.
25. Perriman, R. and Ares, M., Jr. (1998) Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA*, **4**, 1047-1054.
26. Ma, M.Y.-X., McCallum, K., Climie, S.C., Kuperman, R., Lin, W.C., Sumner-Smith, M. and Barnett, R.W. (1993) Design and synthesis of RNA miniduplexes via a synthetic linker approach. 2. Generation of covalently closed, double-stranded cyclic HIV-1 TAR RNA analogs with high Tat-binding affinity. *Nucleic Acids Res*, **21**, 2585-2589.
27. Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J. and Kjems, J. (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J*, **30**, 4414-4422.
28. Abe, N., Abe, H., Nagai, C., Harada, M., Hatakeyama, H., Harashima, H., Ohshiro, T., Nishihara, M., Furukawa, K., Maeda, M. et al. (2011) Synthesis, structure, and biological activity of dumbbell-shaped nanocircular RNAs for RNA interference. *Bioconjug Chem*, **22**, 2082-2092.

29. Lundin, P., Moreno, P.M., Smith, C.I. and Andaloussi, S.E. (2011) Circular RNA interference effector molecules (WO10084371). *Expert Opin Ther Pat*, **21**, 115-119.
30. Peebles, C.L., Perlman, P.S., Mecklenburg, K.L., Petrillo, M.L., Tabor, J.H., Jarrell, K.A. and Cheng, H.L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213-223.
31. van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F. and Grivell, L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, **44**, 225-234.
32. Murray, H.L., Mikheeva, S., Coljee, V.W., Turczyk, B.M., Donahue, W.F., Bar-Shalom, A. and Jarrell, K.A. (2001) Excision of group II introns as circles. *Mol Cell*, **8**, 201-211.
33. Zaphiropoulos, P.G. (1996) Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc Natl Acad Sci U S A*, **93**, 6536-6541.
34. Zhang, Y., Zhang, X.O., Chen, T., Xiang, J.F., Yin, Q.F., Xing, Y.H., Zhu, S., Yang, L. and Chen, L.L. (2013) Circular intronic long noncoding RNAs. *Mol Cell*, **51**, 792-806.
35. Grabowski, P.J., Zaugg, A.J. and Cech, T.R. (1981) The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of tetrahymena. *Cell*, **23**, 467-476.
36. Kruger, K., Grabowski, P.J., Zaugg, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, **31**, 147-157.
37. Sontheimer, E.J., Gordon, P.M. and Piccirilli, J.A. (1999) Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev*, **13**, 1729-1741.
38. Copertino, D.W., Shigeoka, S. and Hallick, R.B. (1992) Chloroplast group III twintron excision utilizing multiple 5'- and 3'-splice sites. *EMBO J*, **11**, 5041-5050.
39. Buzayan, J.M., Hampel, A. and Bruening, G. (1986) Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA. *Nucleic Acids Res*, **14**, 9729-9743.
40. Flores, R., Gas, M.E., Molina-Serrano, D., Nohales, M.A., Carbonell, A., Gago, S., De la Pena, M. and Daros, J.A. (2009) Viroid replication: rolling-circles, enzymes and ribozymes. *Viruses*, **1**, 317-334.
41. Flores, R., Grubb, D., Elleuch, A., Nohales, M.A., Delgado, S. and Gago, S. (2011) Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. *RNA Biol*, **8**, 200-206.
42. Agris, C.H., Nemeroff, M.E. and Krug, R.M. (1989) A block in mammalian splicing occurring after formation of large complexes containing U1, U2, U4, U5, and U6 small nuclear ribonucleoproteins. *Mol Cell Biol*, **9**, 259-267.
43. Berget, S.M. and Robberson, B.L. (1986) U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. *Cell*, **46**, 691-696.
44. Chabot, B. and Steitz, J.A. (1987) Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. *Mol Cell Biol*, **7**, 281-293.
45. Pikielny, C.W. and Rosbash, M. (1986) Specific small nuclear RNAs are associated with yeast spliceosomes. *Cell*, **45**, 869-877.
46. Black, D.L. and Pinto, A.L. (1989) U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol Cell Biol*, **9**, 3350-3359.
47. Burgess, S., Couto, J.R. and Guthrie, C. (1990) A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell*, **60**, 705-717.
48. Kim, S.H. and Lin, R.J. (1993) Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. *Proc Natl Acad Sci U S A*, **90**, 888-892.
49. Reed, R. (1989) The organization of 3' splice-site sequences in mammalian introns. *Genes Dev*, **3**, 2113-2123.

50. Freyer, G.A., O'Brien, J.P. and Hurwitz, J. (1989) Alterations in the polypyrimidine sequence affect the in vitro splicing reactions catalyzed by HeLa cell-free preparations. *J Biol Chem*, **264**, 14631-14637.
51. Cristea, S., Gregory, P.D., Urnov, F.D. and Cost, G.J. (2011) Dissection of splicing regulation at an endogenous locus by zinc-finger nuclease-mediated gene editing. *PLoS One*, **6**, e16961.
52. Fu, X.D., Katz, R.A., Skalka, A.M. and Maniatis, T. (1991) The role of branchpoint and 3'-exon sequences in the control of balanced splicing of avian retrovirus RNA. *Genes Dev*, **5**, 211-220.
53. Reed, R. and Maniatis, T. (1988) The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev*, **2**, 1268-1276.
54. Teigelkamp, S., Newman, A.J. and Beggs, J.D. (1995) Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J*, **14**, 2602-2612.
55. Nilsen, T.W. (ed.) (1998) *RNA-RNA interactions in nuclear pre-mRNA splicing*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
56. Grainger, R.J. and Beggs, J.D. (2005) Prp8 protein: at the heart of the spliceosome. *RNA*, **11**, 533-557.
57. Moore, M.J. and Sharp, P.A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, **365**, 364-368.
58. Maschhoff, K.L. and Padgett, R.A. (1993) The stereochemical course of the first step of pre-mRNA splicing. *Nucleic Acids Res*, **21**, 5456-5462.
59. Pasman, Z., Been, M.D. and Garcia-Blanco, M.A. (1996) Exon circularization in mammalian nuclear extracts. *RNA*, **2**, 603-610.
60. Cocquerelle, C., Daubersies, P., Majerus, M.A., Kerckaert, J.P. and Bailleul, B. (1992) Splicing with inverted order of exons occurs proximal to large introns. *EMBO J*, **11**, 1095-1098.
61. Surono, A., Takeshima, Y., Wibawa, T., Ikezawa, M., Nonaka, I. and Matsuo, M. (1999) Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Hum Mol Genet*, **8**, 493-500.
62. Schindewolf, C.A. and Domdey, H. (1995) Splicing of a circular yeast pre-mRNA in vitro. *Nucleic Acids Res*, **23**, 1133-1139.
63. Bak, R.O. and Mikkelsen, J.G. (2013) miRNA sponges: soaking up miRNAs for regulation of gene expression. *Wiley Interdiscip Rev RNA*.
64. Dubin, R.A., Kazmi, M.A. and Ostrer, H. (1995) Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene*, **167**, 245-248.
65. Lehmann, K. and Schmidt, U. (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit Rev Biochem Mol Biol*, **38**, 249-303.
66. Christopher, D.A. and Hallick, R.B. (1989) Euglena gracilis chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. *Nucleic Acids Res*, **17**, 7591-7608.
67. Copertino, D.W. and Hallick, R.B. (1993) Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends Biochem Sci*, **18**, 467-471.
68. Kruger, K., Grabowski, P.J., Zaag, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.
69. Cech, T.R. (1990) Self-splicing of group I introns. *Annu Rev Biochem*, **59**, 543-568.
70. Zaag, A.J., Grabowski, P.J. and Cech, T.R. (1983) Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. *Nature*, **301**, 578-583.
71. Zaag, A.J. and Cech, T.R. (1982) The intervening sequence excised from the ribosomal RNA precursor of Tetrahymena contains a 5-terminal guanosine residue not encoded by the DNA. *Nucleic Acids Res*, **10**, 2823-2838.

72. Zaug, A.J., Kent, J.R. and Cech, T.R. (1985) Reactions of the intervening sequence of the Tetrahymena ribosomal ribonucleic acid precursor: pH dependence of cyclization and site-specific hydrolysis. *Biochemistry*, **24**, 6211-6218.
73. Sugimoto, N., Tomka, M., Kierzek, R., Bevilacqua, P.C. and Turner, D.H. (1989) Effects of substrate structure on the kinetics of circle opening reactions of the self-splicing intervening sequence from Tetrahymena thermophila: evidence for substrate and Mg²⁺ binding interactions. *Nucleic Acids Res*, **17**, 355-371.
74. Sullivan, F.X. and Cech, T.R. (1985) Reversibility of cyclization of the Tetrahymena rRNA intervening sequence: implication for the mechanism of splice site choice. *Cell*, **42**, 639-648.
75. Johansen, S. and Vogt, V.M. (1994) An intron in the nuclear ribosomal DNA of Didymium iridis codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell*, **76**, 725-734.
76. Vicens, Q. and Cech, T.R. (2009) A natural ribozyme with 3',5' RNA ligase activity. *Nat Chem Biol*, **5**, 97-99.
77. Been, M.D. and Cech, T.R. (1986) One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity. *Cell*, **47**, 207-216.
78. Inoue, T., Sullivan, F.X. and Cech, T.R. (1986) New reactions of the ribosomal RNA precursor of Tetrahymena and the mechanism of self-splicing. *J Mol Biol*, **189**, 143-165.
79. Hadidi, A. (1986) Relationship of viroids and certain other plant pathogenic nucleic acids to group I and II introns. *Plant Mol Biol*, **7**, 129-142.
80. Dinter-Gottlieb, G. (1986) Viroids and virusoids are related to group I introns. *Proc Natl Acad Sci U S A*, **83**, 6250-6254.
81. Nielsen, H., Fiskaa, T., Birgisdottir, A.B., Haugen, P., Einvik, C. and Johansen, S. (2003) The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA*, **9**, 1464-1475.
82. Zofalova, L., Guo, Y. and Gupta, R. (2000) Junction phosphate is derived from the precursor in the tRNA spliced by the archaeon *Haloferax volcanii* cell extract. *RNA*, **6**, 1019-1030.
83. Dalgaard, J.Z. and Garrett, R.A. (1992) Protein-coding introns from the 23S rRNA-encoding gene form stable circles in the hyperthermophilic archaeon *Pyrobaculum organotrophum*. *Gene*, **121**, 103-110.
84. Haseloff, J. and Gerlach, W.L. (1989) Sequences required for self-catalysed cleavage of the satellite RNA of tobacco ringspot virus. *Gene*, **82**, 43-52.
85. Kiefer, M.C., Daubert, S.D., Schneider, I.R. and Bruening, G. (1982) Multimeric forms of satellite of tobacco ringspot virus RNA. *Virology*, **121**, 262-273.
86. Hughes, S.A., Wedemeyer, H. and Harrison, P.M. (2011) Hepatitis delta virus. *Lancet*, **378**, 73-85.
87. Uhlenbeck, O.C. (1987) A small catalytic oligoribonucleotide. *Nature*, **328**, 596-600.
88. Kuo, M.Y., Goldberg, J., Coates, L., Mason, W., Gerin, J. and Taylor, J. (1988) Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. *J Virol*, **62**, 1855-1861.
89. Wang, K.S., Choo, Q.L., Weiner, A.J., Ou, J.H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L. and Houghton, M. (1986) Structure, sequence and expression of the hepatitis delta (delta) viral genome. *Nature*, **323**, 508-514.
90. Reid, C.E. and Lazinski, D.W. (2000) A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. *Proc Natl Acad Sci U S A*, **97**, 424-429.
91. Jeng, K.S., Su, P.Y. and Lai, M.M. (1996) Hepatitis delta antigens enhance the ribozyme activities of hepatitis delta virus RNA in vivo. *J Virol*, **70**, 4205-4209.
92. Sharmin, L., Kuo, M.Y. and Taylor, J. (1989) Self-ligating RNA sequences on the antigenome of human hepatitis delta virus. *J Virol*, **63**, 1428-1430.

93. Wu, H.N., Lin, Y.J., Lin, F.P., Makino, S., Chang, M.F. and Lai, M.M. (1989) Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci U S A*, **86**, 1831-1835.
94. Wu, H.N. and Lai, M.M. (1989) Reversible cleavage and ligation of hepatitis delta virus RNA. *Science*, **243**, 652-654.
95. Ke, A., Zhou, K., Ding, F., Cate, J.H. and Doudna, J.A. (2004) A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*, **429**, 201-205.
96. Tsagris, E.M., Martinez de Alba, A.E., Gozmanova, M. and Kalantidis, K. (2008) Viroids. *Cell Microbiol*, **10**, 2168-2179.
97. Navarro, B. and Flores, R. (1997) Chrysanthemum chlorotic mottle viroid: unusual structural properties of a subgroup of self-cleaving viroids with hammerhead ribozymes. *Proc Natl Acad Sci U S A*, **94**, 11262-11267.
98. Navarro, J.A. and Flores, R. (2000) Characterization of the initiation sites of both polarity strands of a viroid RNA reveals a motif conserved in sequence and structure. *EMBO J*, **19**, 2662-2670.
99. Lima, M.I., Fonseca, M.E., Flores, R. and Kitajima, E.W. (1994) Detection of avocado sunblotch viroid in chloroplasts of avocado leaves by in situ hybridization. *Arch Virol*, **138**, 385-390.
100. Haseloff, J. and Symons, R.H. (1982) Comparative sequence and structure of viroid-like RNAs of two plant viruses. *Nucleic Acids Res*, **10**, 3681-3691.
101. Lafontaine, D., Beaudry, D., Marquis, P. and Perreault, J.P. (1995) Intra- and intermolecular nonenzymatic ligations occur within transcripts derived from the peach latent mosaic viroid. *Virology*, **212**, 705-709.
102. Cote, F. and Perreault, J.P. (1997) Peach latent mosaic viroid is locked by a 2',5'-phosphodiester bond produced by in vitro self-ligation. *J Mol Biol*, **273**, 533-543.
103. Branch, A.D., Robertson, H.D., Greer, C., Gegenheimer, P., Peebles, C. and Abelson, J. (1982) Cell-free circularization of viroid progeny RNA by an RNA ligase from wheat germ. *Science*, **217**, 1147-1149.
104. Diener, T.O. (1989) Circular RNAs: relics of precellular evolution? *Proc Natl Acad Sci U S A*, **86**, 9370-9374.
105. Dolinnaya, N.G., Sokolova, N.I., Ashirbekova, D.T. and Shabarova, Z.A. (1991) The use of BrCN for assembling modified DNA duplexes and DNA-RNA hybrids; comparison with water-soluble carbodiimide. *Nucleic Acids Res*, **19**, 3067-3072.
106. Kanaya, E. and Yanagawa, H. (1986) Template-directed polymerization of oligoadenylates using cyanogen bromide. *Biochemistry*, **25**, 7423-7430.
107. Wang, S. and Kool, E.T. (1994) Circular RNA oligonucleotides. Synthesis, nucleic acid binding properties, and a comparison with circular DNAs. *Nucleic Acids Res*, **22**, 2326-2333.
108. Carriero, S. and Damha, M.J. (2003) Template-mediated synthesis of lariat RNA and DNA. *J Org Chem*, **68**, 8328-8338.
109. Prakash, G. and Kool, E.T. (1992) Structural effects in the recognition of DNA by circular oligonucleotides. *Journal of the American Chemical Society*, **114**, 3523-3527.
110. Ferris, J.P., Huang, C.H. and Hagan, W.J., Jr. (1989) N-cyanoimidazole and diimidazole imine: water-soluble condensing agents for the formation of the phosphodiester bond. *Nucleosides Nucleotides*, **8**, 407-414.
111. Fedorova, O.A., Gottikh, M.B., Oretskaya, T.S. and Shabarova, Z.A. (1996) Cyanogen Bromide-Induced Chemical Ligation: Mechanism and Optimization of the Reaction Conditions. *Nucleosides Nucleotides Nucleic Acids*, **15**, 1137-1147.
112. Micura, R. (1999) Cyclic Oligoribonucleotides (RNA) by Solid-Phase Synthesis. *Chemistry – A European Journal*, **5**, 2077-2082.
113. Kaufmann, G. and Kallenbach, N.R. (1975) Determination of recognition sites of T4 RNA ligase on the 3'-OH and 5' -P termini of polyribonucleotide chains. *Nature*, **254**, 452-454.

114. Kaufmann, G., Klein, T. and Littauer, U.Z. (1974) T4 RNA ligase: substrate chain length requirements. *FEBS Lett*, **46**, 271-275.
115. Walker, G.C., Uhlenbeck, O.C., Bedows, E. and Gumpert, R.I. (1975) T4-induced RNA ligase joins single-stranded oligoribonucleotides. *Proc Natl Acad Sci U S A*, **72**, 122-126.
116. Last, J.A. and Anderson, W.F. (1976) Purification and properties of bacteriophage T4-induced RNA ligase. *Arch Biochem Biophys*, **174**, 167-176.
117. Silber, R., Malathi, V.G. and Hurwitz, J. (1972) Purification and properties of bacteriophage T4-induced RNA ligase. *Proc Natl Acad Sci U S A*, **69**, 3009-3013.
118. Romaniuk, E., McLaughlin, L.W., Neilson, T. and Romaniuk, P.J. (1982) The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur J Biochem*, **125**, 639-643.
119. Sugino, A., Snoper, T.J. and Cozzarelli, N.R. (1977) Bacteriophage T4 RNA ligase. Reaction intermediates and interaction of substrates. *J Biol Chem*, **252**, 1732-1738.
120. England, T.E. and Uhlenbeck, O.C. (1978) Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry*, **17**, 2069-2076.
121. Nichols, N.M., Tabor, S. and McReynolds, L.A. (2008) RNA ligases. *Curr Protoc Mol Biol*, **Chapter 3**, Unit3 15.
122. Wang, L. and Ruffner, D.E. (1998) Oligoribonucleotide circularization by 'template-mediated' ligation with T4 RNA ligase: synthesis of circular hammerhead ribozymes. *Nucleic Acids Res*, **26**, 2502-2504.
123. Beaudry, D. and Perreault, J.P. (1995) An efficient strategy for the synthesis of circular RNA molecules. *Nucleic Acids Res*, **23**, 3064-3066.
124. Abe, H., Abe, N., Harada, M., Tsuneda, S. and Ito, Y. (2008) Nanocircular RNAs for RNA interference. *Nucleic Acids Symp Ser (Oxf)*, 505-506.
125. Abe, H., Abe, N., Uda, M., Tsuneda, S. and Ito, Y. (2009) Synthetic nanocircular RNA for controlling of gene expression. *Nucleic Acids Symp Ser (Oxf)*, 65-66.
126. Abe, N., Abe, H. and Ito, Y. (2012) Synthesis of dumbbell-shaped cyclic RNAs for RNA interference. *Curr Protoc Nucleic Acid Chem*, **Chapter 16**, Unit 16 14 11-11.
127. Puttaraju, M. and Been, M.D. (1992) Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons. *Nucleic Acids Res*, **20**, 5357-5364.
128. Ford, E. and Ares, M., Jr. (1994) Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proc Natl Acad Sci U S A*, **91**, 3117-3121.
129. Rezaian, M.A. (1999) Synthesis of infectious viroids and other circular RNAs. *Curr Issues Mol Biol*, **1**, 13-20.
130. Wang, Y. and Silverman, S.K. (2006) A general two-step strategy to synthesize lariat RNAs. *RNA*, **12**, 313-321.
131. Wang, Y. and Silverman, S.K. (2005) Efficient one-step synthesis of biologically related lariat RNAs by a deoxyribozyme. *Angew Chem Int Ed Engl*, **44**, 5863-5866.
132. Harris, M.E. and Christian, E.L. (1999) Use of circular permutation and end modification to position photoaffinity probes for analysis of RNA structure. *Methods*, **18**, 51-59.
133. Schindewolf, C., Braun, S. and Domdey, H. (1996) In vitro generation of a circular exon from a linear pre-mRNA transcript. *Nucleic Acids Res*, **24**, 1260-1266.
134. Petkovic, S. and Muller, S. (2013) RNA self-processing: formation of cyclic species and concatemers from a small engineered RNA. *FEBS Lett*, **587**, 2435-2440.
135. Ho, C.K. and Shuman, S. (2002) Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc Natl Acad Sci U S A*, **99**, 12709-12714.
136. Ho, C.K., Wang, L.K., Lima, C.D. and Shuman, S. (2004) Structure and mechanism of RNA ligase. *Structure*, **12**, 327-339.
137. Nandakumar, J., Shuman, S. and Lima, C.D. (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell*, **127**, 71-84.

138. Nandakumar, J., Ho, C.K., Lima, C.D. and Shuman, S. (2004) RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J Biol Chem*, **279**, 31337-31347.
139. Yin, S., Kiong Ho, C., Miller, E.S. and Shuman, S. (2004) Characterization of bacteriophage KVP40 and T4 RNA ligase 2. *Virology*, **319**, 141-151.
140. Konarska, M., Filipowicz, W., Domdey, H. and Gross, H.J. (1981) Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. *Nature*, **293**, 112-116.
141. Konarska, M., Filipowicz, W. and Gross, H.J. (1982) RNA ligation via 2'-phosphomonoester, 3'5'-phosphodiester linkage: requirement of 2',3'-cyclic phosphate termini and involvement of a 5'-hydroxyl polynucleotide kinase. *Proc Natl Acad Sci U S A*, **79**, 1474-1478.
142. Schwartz, R.C., Greer, C.L., Gegenheimer, P. and Abelson, J. (1983) Enzymatic mechanism of an RNA ligase from wheat germ. *J Biol Chem*, **258**, 8374-8383.
143. Gegenheimer, P., Gabius, H.J., Peebles, C.L. and Abelson, J. (1983) An RNA ligase from wheat germ which participates in transfer RNA splicing in vitro. *J Biol Chem*, **258**, 8365-8373.
144. Greer, C.L., Peebles, C.L., Gegenheimer, P. and Abelson, J. (1983) Mechanism of action of a yeast RNA ligase in tRNA splicing. *Cell*, **32**, 537-546.
145. Englert, M. and Beier, H. (2005) Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins. *Nucleic Acids Res*, **33**, 388-399.
146. Brooks, M.A., Meslet-Cladiere, L., Graille, M., Kuhn, J., Blondeau, K., Myllykallio, H. and van Tilbeurgh, H. (2008) The structure of an archaeal homodimeric ligase which has RNA circularization activity. *Protein Sci*, **17**, 1336-1345.
147. Tanaka, N. and Shuman, S. (2011) RtcB is the RNA ligase component of an Escherichia coli RNA repair operon. *J Biol Chem*, **286**, 7727-7731.
148. Chakravarty, A.K., Subbotin, R., Chait, B.T. and Shuman, S. (2012) RNA ligase RtcB splices 3'-phosphate and 5'-OH ends via covalent RtcB-(histidinyl)-GMP and polynucleotide-(3')pp(5')G intermediates. *Proc Natl Acad Sci U S A*, **109**, 6072-6077.
149. Lohman, G.J., Tabor, S. and Nichols, N.M. (2011) DNA ligases. *Curr Protoc Mol Biol*, **Chapter 3**, Unit3 14.
150. Harvey, C.L., Gabriel, T.F., Wilt, E.M. and Richardson, C.C. (1971) Enzymatic breakage and joining of deoxyribonucleic acid. IX. Synthesis and properties of the deoxyribonucleic acid adenylate in the phage T4 ligase reaction. *J Biol Chem*, **246**, 4523-4530.
151. Olivera, B.M., Hall, Z.W., Anraku, Y., Chien, J.R. and Lehman, I.R. (1968) On the mechanism of the polynucleotide joining reaction. *Cold Spring Harb Symp Quant Biol*, **33**, 27-34.
152. Wu, D.Y. and Wallace, R.B. (1989) Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene*, **76**, 245-254.
153. Kurschat, W.C., Muller, J., Wombacher, R. and Helm, M. (2005) Optimizing splinted ligation of highly structured small RNAs. *RNA*, **11**, 1909-1914.
154. Lang, K. and Micura, R. (2008) The preparation of site-specifically modified riboswitch domains as an example for enzymatic ligation of chemically synthesized RNA fragments. *Nat Protoc*, **3**, 1457-1466.
155. Flores, R., Gas, M.E., Molina, D., Hernandez, C. and Daros, J.A. (2008) Analysis of viroid replication. *Methods Mol Biol*, **451**, 167-183.
156. Branch, A.D. and Robertson, H.D. (1984) A replication cycle for viroids and other small infectious RNA's. *Science*, **223**, 450-455.
157. Diegelman, A.M. and Kool, E.T. (1998) Generation of circular RNAs and trans-cleaving catalytic RNAs by rolling transcription of circular DNA oligonucleotides encoding hairpin ribozymes. *Nucleic Acids Res*, **26**, 3235-3241.

158. Feldstein, P.A. and Bruening, G. (1993) Catalytically active geometry in the reversible circularization of 'mini-monomer' RNAs derived from the complementary strand of tobacco ringspot virus satellite RNA. *Nucleic Acids Res*, **21**, 1991-1998.
159. Dallas, A., Balatskaya, S.V., Kuo, T.C., Ilves, H., Vlassov, A.V., Kaspar, R.L., Kisich, K.O., Kazakov, S.A. and Johnston, B.H. (2008) Hairpin ribozyme-antisense RNA constructs can act as molecular Lassos. *Nucleic Acids Res*, **36**, 6752-6766.
160. Kazakov, S.A., Balatskaya, S.V. and Johnston, B.H. (2006) Ligation of the hairpin ribozyme in cis induced by freezing and dehydration. *RNA*, **12**, 446-456.
161. Price, J.V., Engberg, J. and Cech, T.R. (1987) 5' exon requirement for self-splicing of the *Tetrahymena thermophila* pre-ribosomal RNA and identification of a cryptic 5' splice site in the 3' exon. *J Mol Biol*, **196**, 49-60.
162. Liu, Y., Cui, H., Wang, W., Li, L., Wang, Z., Yang, S. and Zhang, X. (2013) Construction of circular miRNA sponges targeting miR-21 or miR-221 and demonstration of their excellent anticancer effects on malignant melanoma cells. *Int J Biochem Cell Biol*, **45**, 2643-2650.
163. Ochi, A., Umekage, S. and Kikuchi, Y. (2009) Non-enzymatic in vitro production of circular hammerhead ribozyme targeting the template region of human telomerase RNA. *Nucleic Acids Symp Ser (Oxf)*, 275-276.
164. Puttaraju, M., Perrotta, A.T. and Been, M.D. (1993) A circular trans-acting hepatitis delta virus ribozyme. *Nucleic Acids Res*, **21**, 4253-4258.
165. Umekage, S., Uehara T., Fujita Y., Suzuki H., Kikuchi, Y. . (2012) In Vivo Circular RNA Expression by the Permuted Intron-Exon Method. *Innovations in Biotechnology*.
166. Mikheeva, S., Hakim-Zargar, M., Carlson, D. and Jarrell, K. (1997) Use of an engineered ribozyme to produce a circular human exon. *Nucleic Acids Res*, **25**, 5085-5094.
167. Jarrell, K.A. (1993) Inverse splicing of a group II intron. *Proc Natl Acad Sci U S A*, **90**, 8624-8627.
168. Ma, W., Yu, C. and Zhang, W. (2013) Circularity and self-cleavage as a strategy for the emergence of a chromosome in the RNA-based protocell. *Biology Direct*.
169. Gopinath, S.C., Misono, T.S., Kawasaki, K., Mizuno, T., Imai, M., Odagiri, T. and Kumar, P.K. (2006) An RNA aptamer that distinguishes between closely related human influenza viruses and inhibits haemagglutinin-mediated membrane fusion. *J Gen Virol*, **87**, 479-487.

Figure Legends

- Fig. 1:** Scheme showing the two transesterifications in spliceosome-mediated pre-mRNA processing. a) Eukaryotic pre-mRNA that is to be spliced, including the polypyrimidine tract (Y_n) and the branchpoint adenosine (bpA); b) RNA folding required for first transesterification, whereby the 2'-OH-group of the bpA is arranged close to the phosphodiester bond linking the 5'-exon with the intron. The nucleophilic attack of the 2'-hydroxyl group of bpA is separately shown in the magnified window; c) nucleophilic attack of the 3'-OH group of the 5'-exon onto the 3'-exon-intron linkage, ligating the two exons, and releasing the snRNPs and the lariat shown in d).
- Fig. 2:** Two models of splice mechanisms leading to regular circRNAs, or to lariats and two-tailed circRNAs: First model (on the left): a) RNA folding resulting in juxtaposed 3'-end of exon 4 and 5'-end of exon 1, leading to combined excision of ligated exons 1 and 4; b) newly formed 5'- and 3'-end of the residual RNA strand converging non-exonic regions for further excision; c) formation of an intermediate RNA circle composed of combined exons 2 (e2), 3 (e3) and intronic RNA; d) final processing of the circRNA intermediate releasing the e2-e3-circle and a lariat. Second model (on the right): e) intronic base pairing is followed by f) intron migration that allows for g) release of exon1 including its flanked non-exonic and partly branched sequences; g) remaining e2 and e3 form a circRNA intermediate that is h) further processed into a regular circle and a lariat analogous to d).
- Fig. 3:** GTP-dependent splicing and subsequent circularization of group-I-introns and 3'-end hydrolysis of the intron: a) Exogenous GTP bound to the intron is used as nucleophile attacking the phosphodiester at the 5'-exon-intron junction and thus becoming linked to the 5'-terminus of the intron. The 5'-exon (E1) is released with a free 3'-hydroxyl group; b) Attack of E1 on the 3'-exon (E2), resulting in joined E1-E2 and the intron containing guanosine at both termini; c) Nucleophilic attack of ω G on a circularization site within the intron, typically 15 or 19 nucleotides from the 5'-terminus, leading to circRNA and a short linear α -GTP-containing RNA fragment; d) Nucleophilic attack of ω G on α G resulting in a full-length intron cycle and release of pyrophosphate (pp); e) Endonucleolytic RNA hydrolysis at the E2-intron junction (3'-end hydrolysis) and release of E2; f) Nucleophilic attack of ω G on the E1-intron junction, generating E1 and a full length intron circle.
- Fig. 4:** Double rolling circle mechanism, occurring in *TobRV*. Note that self-cleavage of the multimeric (-) - strand is mediated by a hammerhead ribozyme, whereas self-cleavage of the multimeric (+) - strand occurs by a hairpin ribozyme. A similar mechanism is employed in the hepatitis- δ -virus replication, as well as in other satellite RNAs and viroids.

- Fig. 5:** Formation of a phosphodiester linkage using BrCN and N-methyl morpholine or imidazole: a) activation of BrCN; b) activation of 3'-phosphate; c) nucleophilic attack of the 5'-hydroxyl group leading to ligation and formation of a circRNA.
- Fig. 6:** Mechanism of ligation (including circularization) by T4 RNA ligase 1 in the presence of magnesium ions as proposed by Sugino et al. (119). E = enzyme, pppA = ATP, EpA = adenylated enzyme, PPi = diphosphate, R = receptor or acceptor, pD = phosphorylated donor sequence, A(5')pp(5')D = reaction intermediate composed of AMP and donor sequence, RpD = product.
- Fig. 7:** Ligation with T4 RNA ligase 1; a) top: with the help of a hairpin structured deoxyoligonucleotide hybridizing in a distance from the reaction site. Bold line: DNA hairpin with DNA-DNA base pairing; thin line: RNA strand base pairing with DNA hairpin template; bottom: with a linear helper DNA (bold line), thin line: RNA; b) T4 RNA ligase 1 mediated circularization of dumbbell shaped RNA structures; c) top: branched RNA with Watson-Crick base paired stem to juxtapose the 5'- and 3'-end for enzymatic ligation, bottom: enzyme-free ligation using a deoxyribozyme (grey).
- Fig. 8:** Splint mediated RNA circularization using T4 DNA ligase: a) Formation of a ligation competent complex; b) ligation competent complex for ligation within a loop.
- Fig. 9:** Schematic presentation of circularization of hairpin ribozyme derived small engineered RNAs. a) Dimerization b) left: self-ligation to a monomeric cycle; right: double self-ligation to a dimeric ring.
- Fig. 10:** Circularization of RNA using the PIE method. Expression vector (relevant part) of the PIE construct containing half introns of the T4 bacteriophage gene *td*, cp = convergent primer sequences to detect linear and circRNA, dp = divergent primer sequences to detect circRNA exclusively. 3'- or 5'-ss denote splice sites, *prom* = promoter sequence of *cytomegalovirus*; a) first transesterification leading to release of the 5'-half intron of the *td* gene, b) second transesterification leading to formation of circRNA and to release of the 3'-half intron.
- Fig. 11:** Scheme of group-II-intron derived inverse splicing, resulting in circRNA: a) RNA fold supporting the initial nucleophilic attack of the branchpoint-2'-hydroxyl group; b) first transesterification leading to a branched intronic sequence; c) second transesterification resulting in released intron sequences and circularized RNA.
- Fig. 12:** Trend of reports on circRNA(except for viroids or hepatitis delta virus records, since the keyword “circular RNA” is permanently used with these species)

Table 1: Comparison of splicing procedures, generating RNA circles

Process	Half time $t_{1/2}$ [min]	Temperature [°C]	Yield [%]
Natural splicing	2	37	87
PIE	30	30	28
PIE	8	42	90

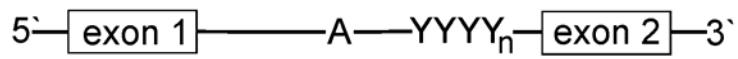
Note that heating to 80 °C and further addition of GTP did not increase circularization efficiency.

Table 2: Circularization strategies for ss and dsRNA *in vitro* and *in vivo*

	ssRNA< 100 nt	ssRNA> 100 nt	dsRNA	Modified nucleobases	RNA
	<i>in vitro</i>				<i>in vivo</i>
Chemically	✓			✓	
T4 RNA ligase 1	✓	✓ dilution			✓
T4 RNA ligase 2			✓		
T4 DNA ligase	✓ with splint	✓ dilution	✓		✓
Rolling circle	✓	✓			✓
Spliceosomal introns					✓
Group-I- introns					✓
Group-II and III-introns	✓	✓			✓
RNA Cyclase	✓	✓			✓
Ribozyme/PIE					
Hammerhead Ribozyme	?	?	✓ partial		✓
Hairpin Ribozyme	✓	✓	✓ partial		✓
HDVR	?	?	?		✓

Figure 1

a)



b)

U2, U5, U6

c)

d)

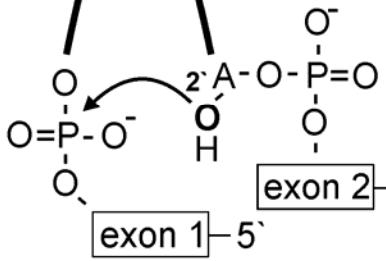


Figure 2

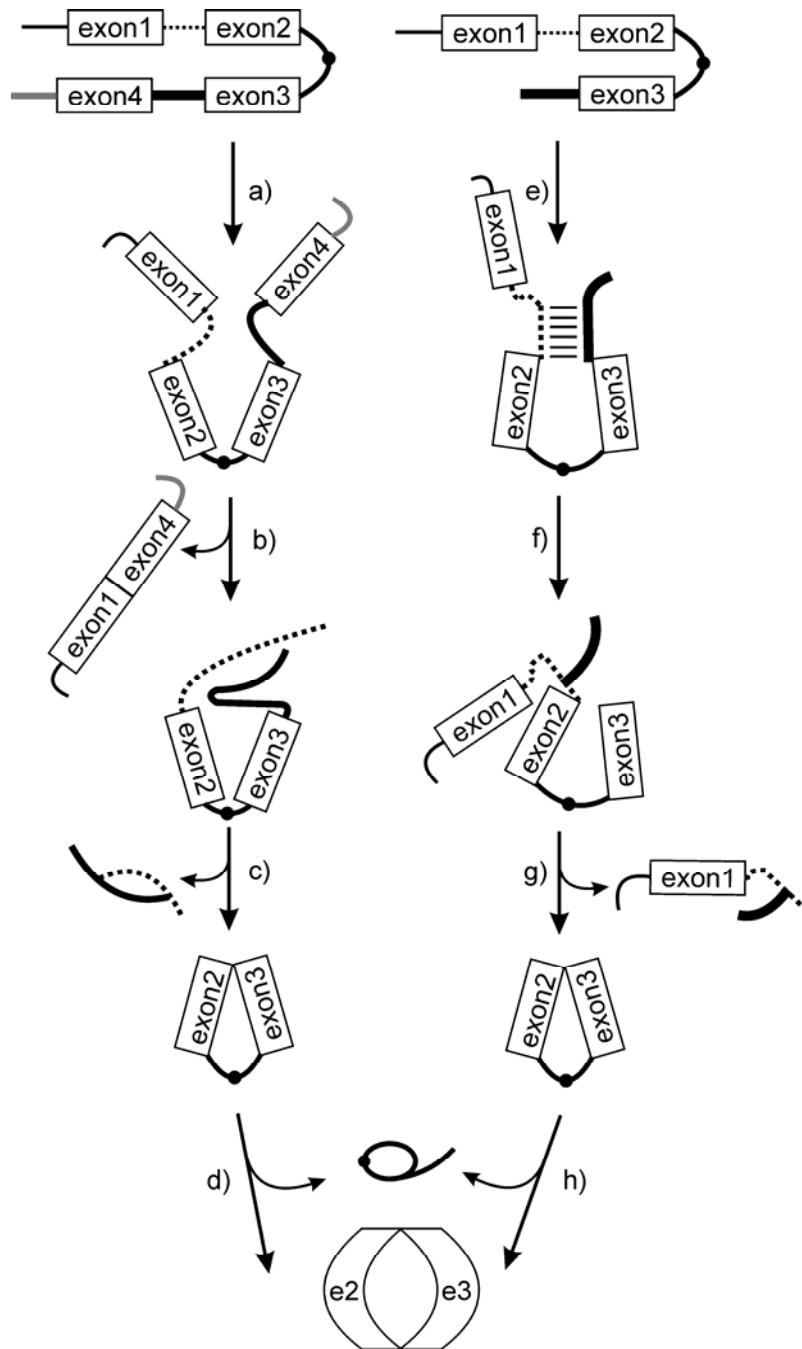


Figure 3

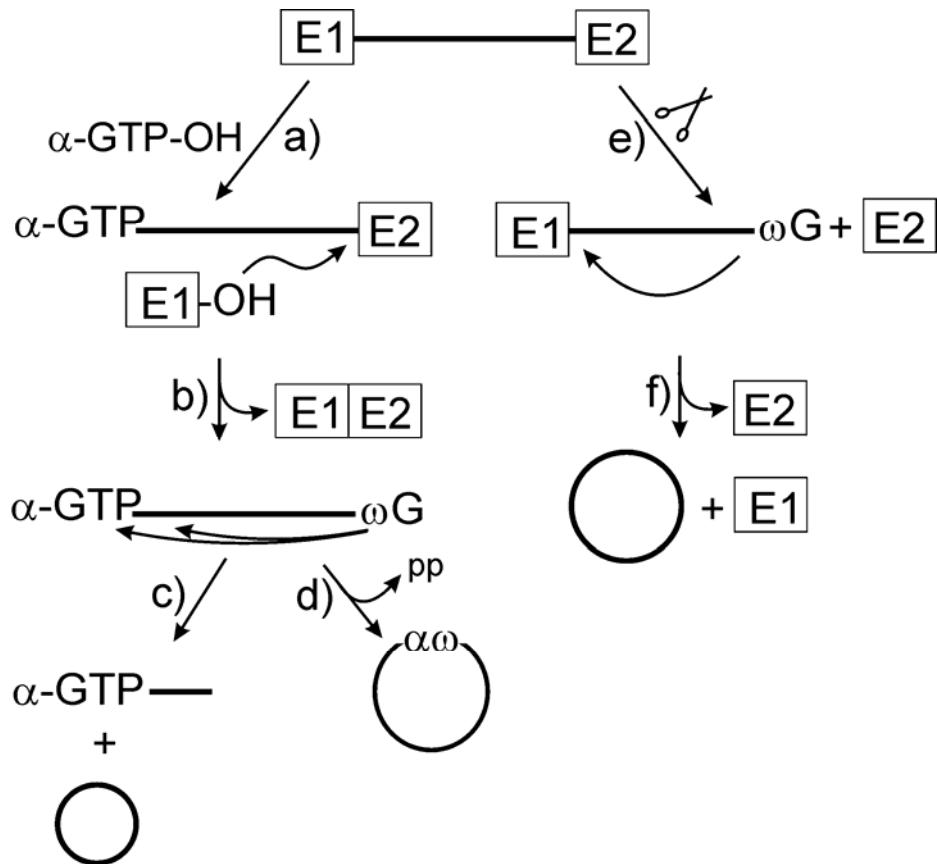


Figure 4

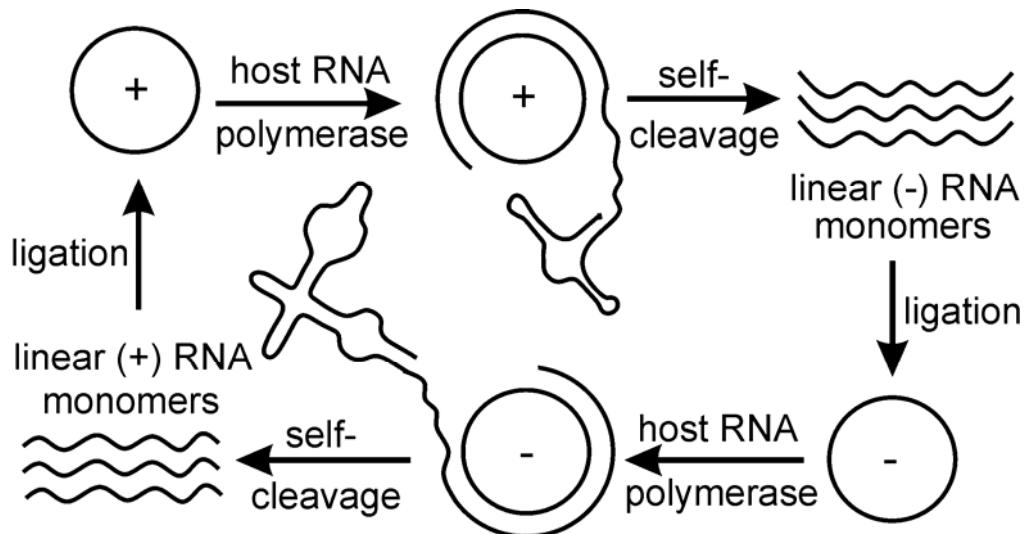
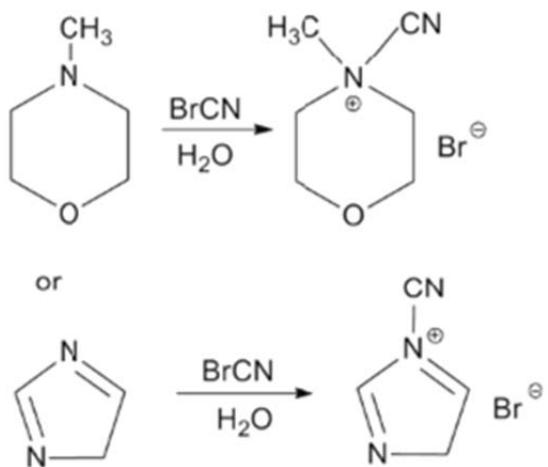


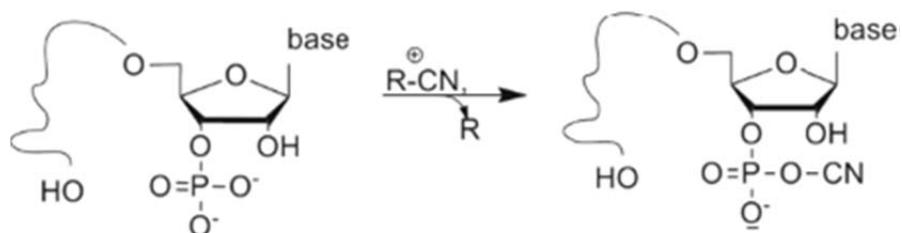
Figure 5

a) activation



NC-R^\oplus = imidazole or morpholino product

b) coupling



c) circularization

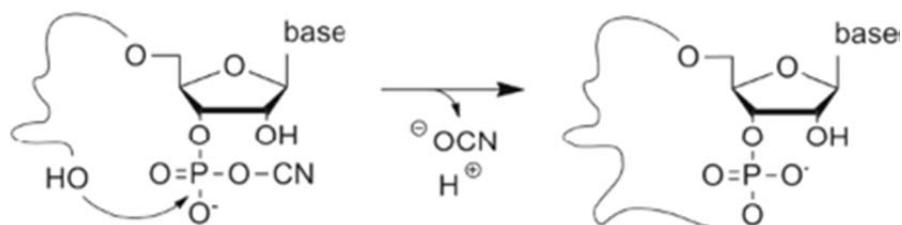


Figure 6

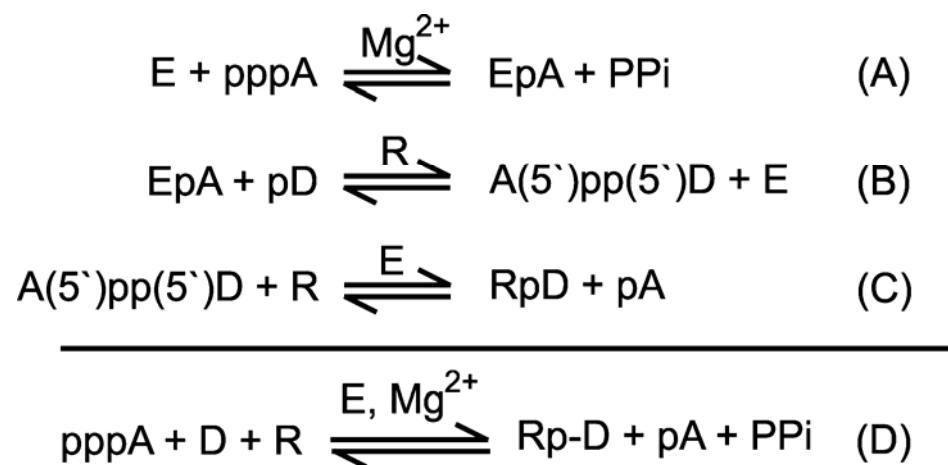


Figure 7

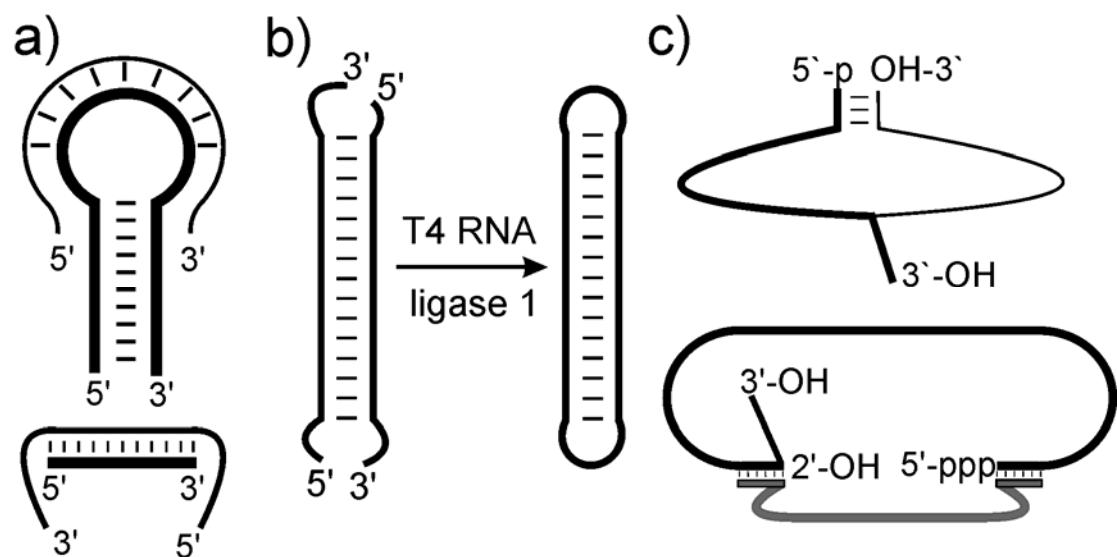


Figure 8

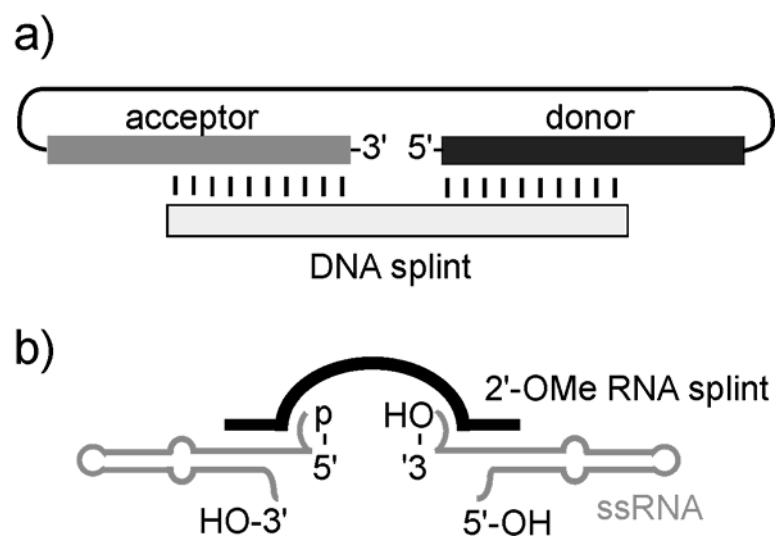


Figure 9

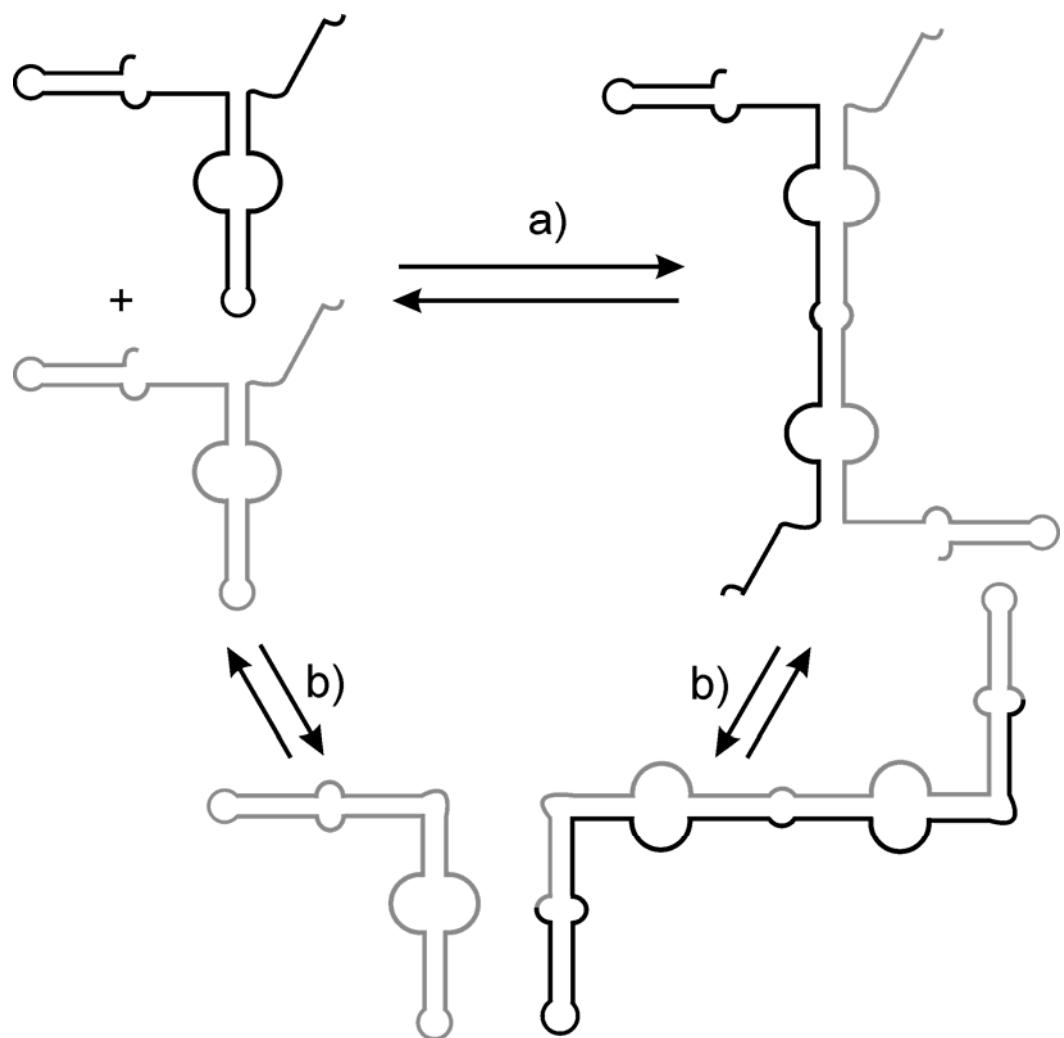


Figure 10

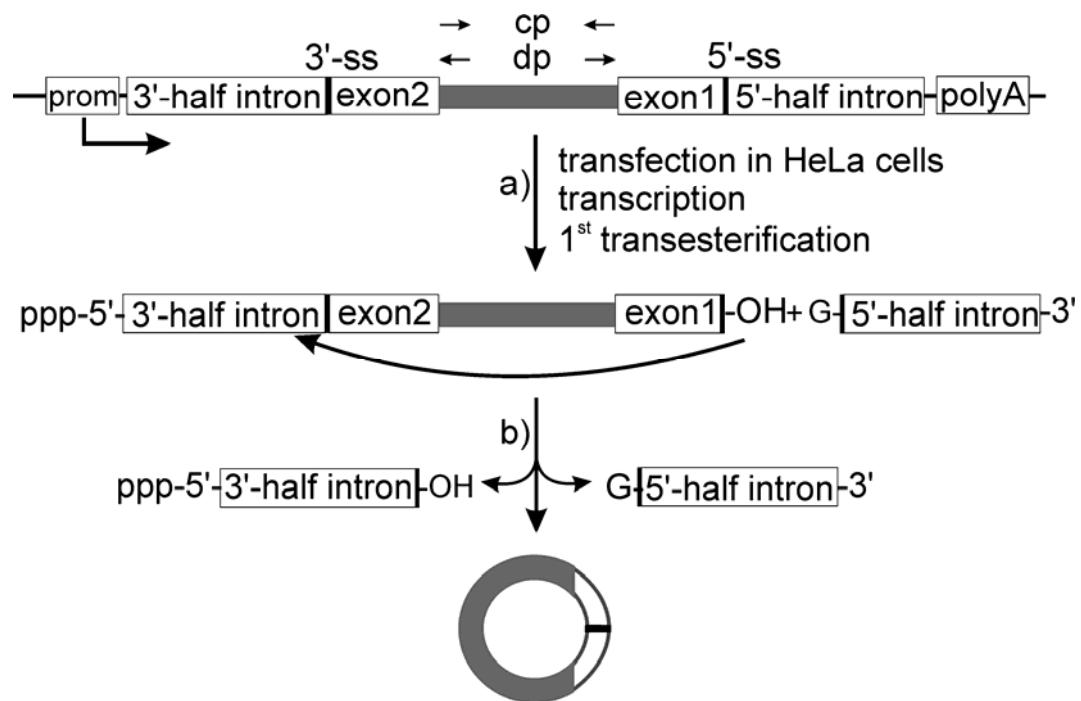


Figure 11

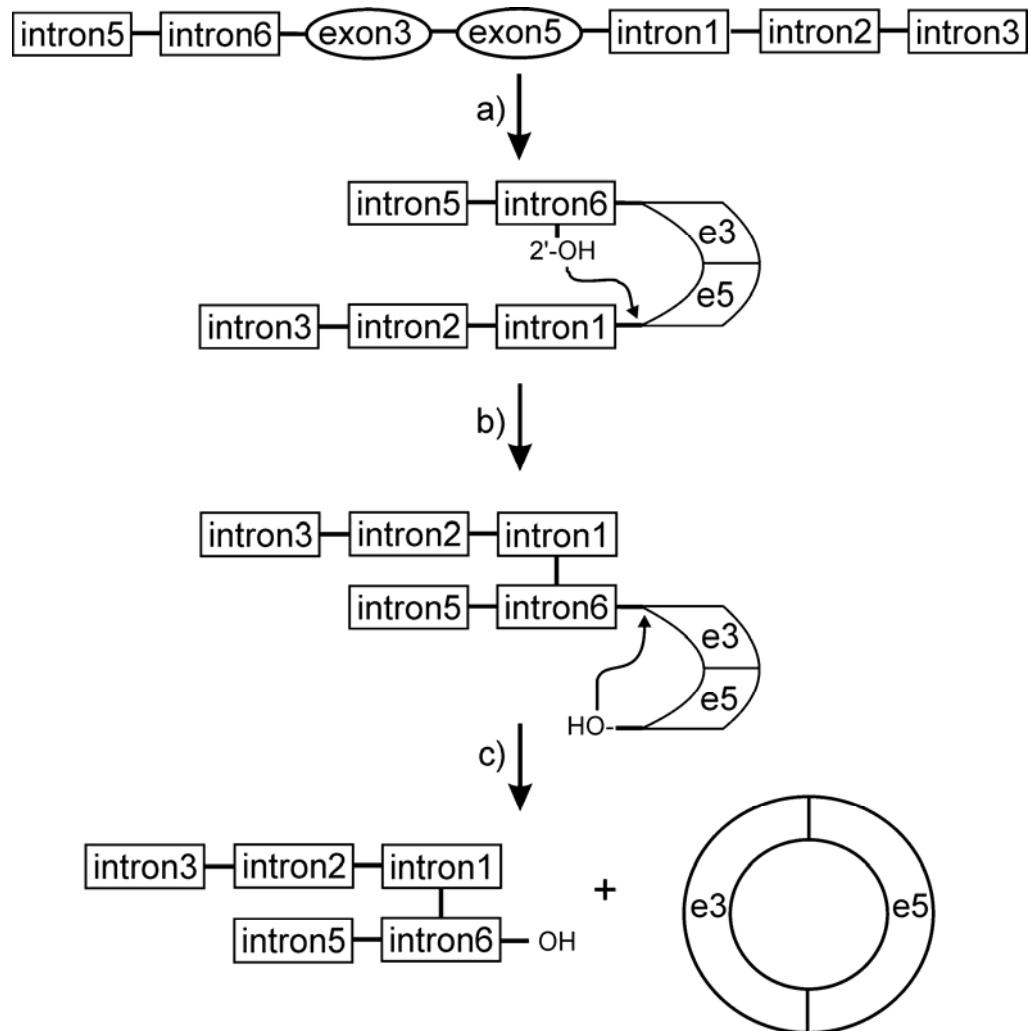
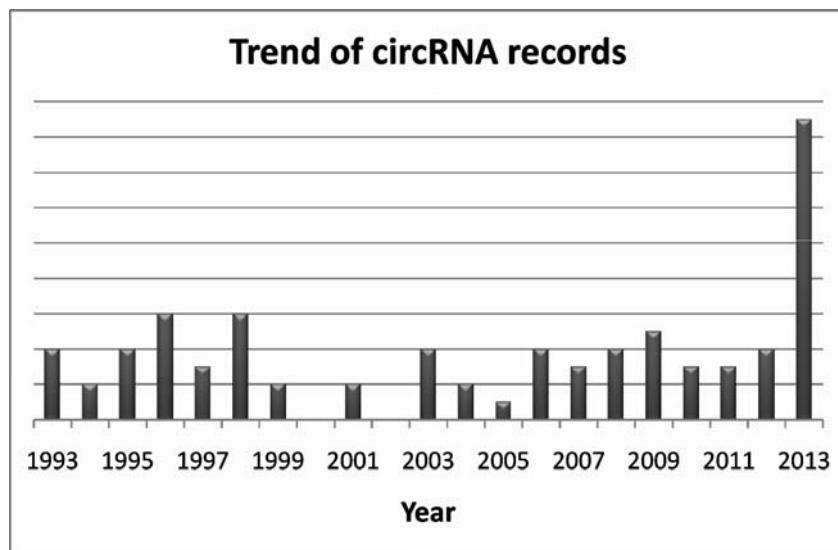


Figure 12



Supplementary Data

Generation of circularRNA *in vivo* and *in vitro*

Sonja Petkovic and Sabine Müller

This PDF file includes:

Supplementary Protocols S1 to S13	p2-p6
References	p7

Protocols

S1: RNA lariat production using cyanogen bromide

Nicked branched RNAs	Each 100 µM
DNA templates	Each 100 µM
MES	0.25 M
MgCl ₂	20 mM
BrCN	1:10 vol of BrCN; 5 M CH ₃ CN

RNAs and DNAs were solved in MES. After addition of MgCl₂, samples were denatured at 95 °C for 10 min, slowly cooled down to room temperature, and then left at 4 °C overnight. The mixture was held on ice for 15 min., then BrCN was added and the following reaction took place on ice within 15 min. Ligation was stopped by precipitation with 1 ml 2% LiClO₄ in acetone (1).

S2: Chemical circularization of RNAs with deoxyuridine at the 3'-end

pre-circRNA	50 µM
aligning DNA	55 µM
Imidazol-HCl (pH = 7.0)	200 mM
NiCl ₂	100 mM
BrCN (solid)	to final concentration of 125 mM

Circularization was conducted for 12 h (2).

S3: One of the first circularization assays with T4 RNA Ligase 1

Tris • Cl, pH 7.5	50 mM
MgCl ₂	10 mM
bovine serum albumine	20 pg/µl
ATP	0.1 mM
RNA [5'- ³² P]-oligo(pA)	10 ⁸ -10 ⁹ cpm/µmol
T4 RNA ligase 1	0.005-0.02 U

Total reaction volume was 25 µl and circularization took place at 37 °C for 30 min. The mixture was then boiled for 2 min (3).

S4: RNA circularization with T4 RNA Ligase 1:

RNA [5'- ³² P]- polyA (0.2 nmol of 5'- ³² P-termini)	7 nmol
TrisxHCl, pH 7.5	5 µM
MgCl ₂	1 µM
Albumin	5 µg
ATP	10 nmol
DTT	0.13 µmol
T4 RNA ligase 1	0.005-0.02 U

Total reaction volume was 100 µl and circularization took place at 38 °C for 30 min. Reaction was stopped by boiling for 2 min (3).

S5: Circularization of single-stranded oligoribo- or oligodeoxyribonucleotides with T4 RNA ligase 1:

RNA or DNA with 3'-hydroxyl terminus	6 pmol
TrisxHCl, pH 7.8	50 mM
MgCl ₂	10 mM
DTT	10 mM
ATP	1 mM
hexamine cobalt chloride	1 mM
PEG 8000	25% (w/v)
T4 RNA ligase 1	20 U

Reaction was carried out for 16 hr at 22 °C in a total reaction volume of 10 µl and stopped by addition of 40 µl of 10 mM TrisxHCl, pH 8.0; 2.5 mM EDTA (4).

S6: Circularization of linear hammerhead ribozymes with T4 RNA ligase 1

RNA	0.03-100 µM
Helper DNA	0.036-120 µM
MgCl ₂	10 mM
T4 RNA ligase 1	0.7 units/µL

RNA mixed with MgCl₂ was denatured (90 °C, 3 min) and incubated in T4 RNA ligase buffer at 16 °C for at least 1 h, before addition of T4 RNA ligase. Reaction was left to proceed at 16 °C overnight (5).

S7: Double-circularization of dumbbell shaped RNA with RNA ligase 1

5'-phosphorylated RNA	2 µM
ATP	1 mM
PEG 6000	25%
bovine serum albumin (BSA)	0.006%
Tris-HCl	50 mM (pH 7.5)
MgCl ₂	10 mM
DTT	10 mM
T4 RNA ligase 1	0.05-0.4 units/µL

Reaction took place at room temperature overnight (6).

S8: Circularization of *in vitro* transcribed single stranded RNA

a) Dephosphorylation:

5'-triphosphate carrying RNA	100 ng
Tris-HCl	10 mM (pH 8.3)
MgCl ₂	1 mM
ZnCl ₂	1 mM
calf intestinal phosphatase	1 unit

Reaction took place at 37°C for 30 min, and was stopped by phenol extraction followed by ethanol precipitation.

b) Kinase reaction:

dephosphorylated RNA	20 µl
Tris-HCl	50 mM (pH 7.6)
MgCl ₂	10 mM
ATP	100 µM
DTT	10 mM
polynucleotide kinase	5 units

Reaction proceeded at 37 °C for 30 min, and was stopped by heating to 65°C for 10 min.

c) Circularization:

RNA	2-3 µg/ml
Hepes	50 mM (pH 8.3)
MgCl ₂	10 mM
ATP	1 mM
DTT	10 mM
DMSO	10%
T4 RNA ligase 1	100units/ml

Circularization took place at 12 °C over 4 h (7).

S9: Circularization of branched single-stranded RNA with T4 RNA ligase 1:

5'-phosphorylated RNA	1 µM
HEPES	3,75 mM (pH=7.5)
NaCl	15 mM
EDTA	0.1 mM
+ HEPES to final concentration of	5 mM
ATP	50 µM
MgCl ₂	10 mM
DTT	10 mM
T4 RNA ligase 1	1 U/µL

RNA, HEPES, NaCl and EDTA was incubated and boiled at 95 °C for 3 min, followed by chilling on ice for 5 min. HEPES concentration was increased to 5 mM and ATP, MgCl₂, DTT and ligase were added as stated. Circularization was conducted at 37 °C for 30 min and stopped using 80 % formamide in 1x TBE, 50 mM EDTA and 0.025% each bromphenol blue and xyclene cyanol (8).

S10: Ligation of a nick in a double-stranded substrate with T4 RNA ligase 2:

dsRNA nicks with adjacent 3'-OH and 5'-p termini	10 to 20 pmol
Tris • Cl, pH 7.5	50 mM
MgCl ₂	2 mM
DTT	1 mM
ATP	400 µM
T4 RNA ligase 2	1 U

Total reaction volume was 20 µl. Reaction proceeded for 30 min at 37 °C and was stopped by adding 2 µl of 0.5 M EDTA. (9)

S11: Exemplary protocol of RNA circularization using Pap1020:

labeled RNA substrate	1 pmol
Tris • Cl, pH 6.5	50 mM
MgCl ₂	5 mM
DTT	5 mM
ATP	5 µM
Pap1020	2 pmol U

Incubate at 75 °C for 30 min (10).

S12: Exemplary protocol of RNA circularization with RtcB in a 10 µl reaction:

RNA substrate	0.1 µM
Tris • Cl, pH 8.0	50 mM
MnCl ₂	2 mM
GTP	6.25 µM
Pap1020	1 µM

Incubation at 37 °C for 30 min. Reaction was stopped using 10 µl 90% (vol/vol) formamide and 50 mM EDTA (11)

S13: Splinted RNA ligation using T4 DNA ligase

Plasmid DNAs were linearized with SmaI, followed by GMP primed *in vitro* transcription with T7 RNA polymerase [10:1 ratio of GMP to GTP]. RNA transcripts and DNA splint were mixed together and denatured at 90 °C. Annealing took place by slow cooling to room temperature in a solution containing the following components within 60 min:

Tris-HCl, pH 7.5	10 mM
NaCl	100 mM
EDTA	0.1 mM
ATP	5 µM
T4 DNA ligase	100 units/ng transcript

Ligation buffer and T4 DNA ligase were added. Circularization proceeded at room temperature for 8 to 16 hours (12)

References

1. Carriero, S. and Damha, M.J. (2003) Template-mediated synthesis of lariat RNA and DNA. *J Org Chem*, **68**, 8328-8338.
2. Wang, S. and Kool, E.T. (1994) Circular RNA oligonucleotides. Synthesis, nucleic acid binding properties, and a comparison with circular DNAs. *Nucleic Acids Res*, **22**, 2326-2333.
3. Kaufmann, G., Klein, T. and Littauer, U.Z. (1974) T4 RNA ligase: substrate chain length requirements. *FEBS Lett*, **46**, 271-275.
4. Nichols, N.M., Tabor, S. and McReynolds, L.A. (2008) RNA ligases. *Curr Protoc Mol Biol, Chapter 3, Unit3 15*.
5. Wang, L. and Ruffner, D.E. (1998) Oligoribonucleotide circularization by 'template-mediated' ligation with T4 RNA ligase: synthesis of circular hammerhead ribozymes. *Nucleic Acids Res*, **26**, 2502-2504.
6. Abe, N., Abe, H., Nagai, C., Harada, M., Hatakeyama, H., Harashima, H., Ohshiro, T., Nishihara, M., Furukawa, K., Maeda, M. et al. (2011) Synthesis, structure, and biological activity of dumbbell-shaped nanocircular RNAs for RNA interference. *Bioconjug Chem*, **22**, 2082-2092.
7. Rezaian, M.A. (1999) Synthesis of infectious viroids and other circular RNAs. *Curr Issues Mol Biol*, **1**, 13-20.
8. Wang, Y. and Silverman, S.K. (2006) A general two-step strategy to synthesize lariat RNAs. *RNA*, **12**, 313-321.
9. Nandakumar, J., Ho, C.K., Lima, C.D. and Shuman, S. (2004) RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J Biol Chem*, **279**, 31337-31347.
10. Brooks, M.A., Meslet-Cladiere, L., Graille, M., Kuhn, J., Blondeau, K., Myllykallio, H. and van Tilbeurgh, H. (2008) The structure of an archaeal homodimeric ligase which has RNA circularization activity. *Protein Sci*, **17**, 1336-1345.
11. Chakravarty, A.K., Subbotin, R., Chait, B.T. and Shuman, S. (2012) RNA ligase RtcB splices 3'-phosphate and 5'-OH ends via covalent RtcB-(histidinyl)-GMP and polynucleotide-(3')pp(5')G intermediates. *Proc Natl Acad Sci U S A*, **109**, 6072-6077.
12. Chen, C.Y. and Sarnow, P. (1995) Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science*, **268**, 415-417.

Evolution and Characterization of Aptamers targeting Platelet Factor 4

Thomas Marschall¹, Sven Brandt², Martin Kreimann², Sonja Petkovic³, Jennifer Frommer³, Andreas Greinacher⁴, Mihaela Delcea², Sabine Müller³

¹Klinik für Innere Medizin A, Forschungscluster 3, Fleischmannstr. 41 Universitätsklinikum, 17487 Greifswald, ²Innovation Center - Humoral Immune Responses in Cardiovascular Disorders, Fleischmannstr. 42-44, 17489 Greifswald, ³dept. Bioorganic Chemistry, Institut für Biochemie Felix-Hausdorffstr. 4D-17487 Greifswald, ⁴Institut für Immunologie und Transfusionsmedizin, Ernst-Moritz-Arndt-Universität, D-17487 Greifswald, Germany

Abstract:

Platelet factor 4 (PF4) is a small, basic chemokine involved in heparin induced thrombocytopenia (HIT), an antibody-mediated response to heparin. Up to now, its biological function remains obscure. We have used the SELEX process (systematic evolution of ligands by exponential enrichment) to identify two classes of RNA-Aptamers with different binding characteristics. Aptamers of class I are prone to form large, multimolecular complexes with PF4, whereas members of class II bind PF4 specifically. We have investigated the binding properties of class I Aptamers by gel shift analysis and photon correlation spectroscopy. Aptamers that bind PF4 specifically were investigated for their structural effects on and their binding affinity to PF4 by CD spectroscopy and AFM.

Introduction:

PF4 is a small, homotetrameric chemokine usually stored in platelets. Upon injuries of blood vessels, e.g. surgery, those thrombocytes release their content into the blood stream. Once released, PF4 is involved in several processes, but its defined biological role remains obscure.

Beyond its role as cytokine, PF4 is a chemoattractant to immune cells, such as white blood cells (Deuel *et al.* 1981). Additionally, PF4 was shown to have the same effects on fibroblasts (Senior *et al.* 1983). As a result, PF4 mediates the migration of fibroblasts to the point of vascular lesion, where they form collagen and participate actively in the repair of blood vessels.

Latest findings suggest that PF4 also plays a vital role in the resistance to bacteria (Krauel *et al.* 2012). Upon binding to PF4, bacteria will be marked for elimination by the immune system. In this context, lipid A has been identified as binding site for PF4 on the surface of gram-positive bacteria.

Furthermore, PF4 has been linked to pathological events such as heparin induced thrombocytopenia (HIT) (Greinacher *et al.* 1994). Due to its highly positively charged surface, PF4 is capable of binding various polyanionic molecules (Kaplan *et al.* 1979;

Rucinski *et al.* 1979; Deuel *et al.* 1977). Upon binding to heparin, its natural target, PF4 neutralizes the anticoagulatory effects of heparin. In some cases, however, the binding of PF4 to heparin leads to the formation of macromolecular complexes that are recognized by IgG antibodies. The antibody binding induces the formation of immune complexes that bind to the surface of thrombocytes, which leads to the release of more PF4. As a result of the increased PF4 concentration, the cytokine binds to heparin-like molecules on the surface of endothelial cells. These complexes are recognized by IgG antibodies, causing severe damage of endothelial cells (Greinacher *et al.* 2006; Rauova *et al.* 2005; Warkentin and Greinacher 2004). Since Aptamers have emerged as a potent class of therapeutics, the selection of Aptamers to PF4 is of tremendous interest.

Here we report the selection of two classes of Aptamers. In a first selection, we have selected Aptamers that are especially prone to form complexes, similar to those formed between PF4 and heparin. If those complexes consisting of PF4 and RNA also show immunogenic effects, care has to be taken when applying Aptamers as potential therapeutics. Aptamers that share common structural features with the selected Aptamers to PF4 might not be suitable for clinical usage due to cross reactivity with PF4. Thus, structural comparisons with the here selected Aptamers might facilitate the development and optimization for future therapeutic Aptamers. On the other hand, Aptamers that bind PF4 specifically are useful to further investigate the cellular functions of PF4. Binding to an Aptamer could prevent PF4 from fulfilling its cellular functions, and thus will help to reveal the biological role of PF4. Recently we have shown that Aptamers exert their immunogenicity in a concentration dependent manner (Jaax *et al.* 2013). If a specific Aptamer could be applied in low concentrations, it might be able to prevent the formation of heparin-PF4 complexes and thus could be a useful antidote.

Materials and Methods:

Platelet factor 4 purified from human blood was purchased from ChromaTec GmbH (Greifswald, Germany) as lyophilized powder. The protein was resuspended in double distilled water to a concentration of 1 mg/ml and stored in aliquots at 4 °C. DNA-oligonucleotides were purchased from Biomers (Ulm, Germany)

SELEX Protocols

The DNA template library (5'-GGGAGAGCCATACCTGAC-N50-GTCAGTCGTCAGGATCCGTG-3') was chemically synthesized. Oligonucleotide primers (5'-AGCGAATTCTAATACGACTCACTATAGGGAGAGCCATACCTGAC-3' (PM1) and 5'-CACGGATCCTGACGACTGAC-3' (PM2)) were also chemically prepared and used in template amplification and reverse transcription. Double stranded DNA templates were obtained by PCR using a mixture of Taq- and Vent-

Polymerase (New England Biolabs, Beverly, USA). Approximately 500 pmol of the library was transcribed using T7-RNA polymerase. RNAs were purified from denaturing polyacrylamid gels (7 M urea) by excising the respective bands, crushing the gel pieces and submerging the pieces in 0.3 M NaOAc (pH 5-7) for several hours.

Selection of complex-forming Aptamers

500 pmol of RNA, kept constant during the selection, was denatured at 90°C for 5 min. Subsequently the RNA was cooled to the given temperature (on ice to 37°C) but was left at 70°C and 50°C, respectively, for 10 min in between to allow proper folding. After leaving the RNA at the respective reaction temperature (see supplementary data for exact conditions) for another 15 min it was mixed with PF4 and left for another 45 min. Both, the RNA and PF4 were solved in 25 mM Tris-acetate pH 7.4 supplemented with 5 mM MgCl₂. After incubation, the mixture was transferred to a spin tube (Millipore), consisting of polyethersulfone, with an exclusion limit of 50 kDa. To separate unbound sequences, the mixtures were centrifuged for 20 min at 5000 rpm at 4°C. After centrifugation, remaining complexes were disrupted by the addition of 1 M NaCl solution and eluted by centrifugation. Subsequently, the salt was removed by gel filtration using Sephadex G25 (Sigma Aldrich), and PF4 was digested by adding Proteinase K (Thermo Scientific) in 25 mM Tris-HCl supplemented with 1 mM CaCl₂, pH 7.4. Following isolation, the RNA was annealed to primer PM2 and revers transcribed using SuperScript III reverse transcriptase (Invitrogen). The resultant cDNA was amplified with PCR in 10 cycles with primers PM1 and PM2. Transcription of the PCR product gave a new library enriched for sequences binding to PF4. Heparin (Sigma Aldrich) was used as a competitor and added after the first centrifugation step to disrupt complexes formed between RNA and PF4. After addition of heparin, the following procedures were as described. However, an additional digestion step with Heparinase I (Sigma Aldrich) had to be included, prior to the addition of Proteinase K. The final selection round was performed with spin tubes consisting of cellulose to deplete the pool of filter binding RNAs.

After 13 of selection, the final pool was cloned using *E.coli* strain DH5α. Therefore, the final DNA library was cloned into the pUC18 vector and transformed into the aforementioned *E.coli* strain. Blue white screening was used to determine which colony carried the modified vector. White colonies were chosen and sent out for sequencing (GATC Biotech, Konstanz, Germany).

Selection of specific Aptamers

For the selection of Aptamers specifically binding to PF4, a buffer system was used to maintain physiological conditions with relevant cations in physiological concentration. Therefore, the RNA and PF4 were solved in 50 mM Tris-HCl pH 7.4 supplemented with 150 mM NaCl; 5 mM KCl; 2.5 CaCl₂ and 1 mM MgCl₂. Selection

was started with the library which was received after the third round of the first selection, to avoid additional pre-selection. The RNA was denatured at 90°C for 5 min. Subsequently the RNA was cooled to 37°C but was left at 70°C and 50°C, respectively, for 10 min in between to allow proper folding. After leaving the RNA at 37°C for another 15 min it was mixed with PF4 and heparin (starting from round 5) and left for another 45 min. 0.45 µm nitrocellulose filters (Invitrogen) were used to remove unbound sequences. After washing, filters were transferred into 20 mM Tris-HCl, 50 mM NaCl and 4 mM CaCl₂, pH 7.4 for digestion with Heparinase I and Proteinase K. After retrieval RNA was revers transcribed and the cDNA amplified as described for the complex forming Aptamers.

After eight rounds of selection the final pool was cloned using *E.coli* strain TG1. All other steps were carried out identically as described for the previous selection.

Gel Shift Analysis

To evaluate the binding properties of both complex forming and specific Aptamers gel shift assays were used. Therefore, an equal amount of the respective Aptamer and PF4 were incubated in a total volume of 20 µl and subsequently applied to a native 2% agarose gel which contained ethidium bromide. For complex forming Aptamers the assays were conducted at NaCl concentrations of 600 mM, 800 mM and 1 M in 25 mM Tris-acetate, pH 7.4 supplemented with 5 mM MgCl₂.

In the case of specific Aptamers, 10% native polyacrylamide gels were used. To visualize the protein containing bands, the protein sensitive dye SYPRO Orange (Sigma Aldrich) was used. To allow staining, the gel was soaked in 0.05% SDS solution for approximately 30 min and rinsed with water. Afterwards, it was transferred into a solution of SYPRO Orange in 7.5% acetic acid and left there for another 30 min. Gels were visualized utilizing an Infinity photo system (Vilber Lourmat).

Photon Correlation Spectroscopy

PF4 was resuspended in 25 mM Tris-acetate, pH 7.4 supplemented with 5 mM MgCl₂ containing NaCl in the necessary concentration (0.6 - 1 M) and filtered afterwards. 50 pmol of each the Aptamers and PF4 were incubated for 45 min at 37°C. Amount of free PF4 at each NaCl concentration was determined with PCS (Zetasizer 3000HS, Malvern Instruments, UK) by determining the mean hydrodynamic diameter (MHD) value

CD-Spectroscopy

CD-spectroscopy was used to determine K_D-values for the Aptamers. Changes in the secondary structure of PF4 upon complex formation with RNA Aptamers were

investigated by Circular Dichroism (CD) spectroscopy using a Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK). Far UV spectra (200-260 nm) were recorded for different molar ratios of PF4 and Aptamers (indicated in the text) at 20°C using a 5 mm pathlength cuvette (110-QS, Hellma, Mühlheim, Germany).

CD-measurements were started with pure PF4 (1.25×10^{-6} M) in 1x PBS buffer, followed by recording spectra with subsequently increased RNA concentration. Baseline spectra of both PBS and RNA at the corresponding concentrations were subtracted and resulting spectra were corrected for concentration, number of amino acids and pathlength to obtain mean residue delta epsilon (MRDE) values. To obtain information of the secondary structure content of the protein, deconvolution of the CD spectra was conducted at each concentration step. Different molar ratios were created by subsequently increasing the Aptamer concentration ($0-6.3 \times 10^{-7}$ M for SX1 and $0-3.1 \times 10^{-7}$ M for SX4) in the cuvette.

AFM measurements

RNA preparation

RNA (50.40 and SX4) lyophilisates were solved in millipore water. Sodium chloride (final concentration 600 mM) was added, RNAs were denatured (90 °C, 2-5 min) and slow cooling to 37 °C for proper RNA folding was followed by „reaction“ at 37 °C for 30 min. Sample was dropped on freshly cleaved mica, adsorption took place within 2 min and was followed by washing of mica surface with millipore for 30 s and drying in a laminar flow for ca. 15 min. Fresh or maximal one month old samples were imaged as described in section AFM imaging.

PF4 preparation

PF4 in PBS buffer was mixed with sodium chloride (final concentration 600 mM) and folding was allowed to take place at 37 °C for 30 min. Sample was dropped on freshly cleaved mica, adsorption took place within 2 min and was followed by washing of mica surface with millipore for 30 s and drying in a laminar flow for ca. 15 min. Fresh or maximal one month old samples were imaged as described in section AFM imaging.

PF4-RNA complexes

RNAs were denatured (90 °C, 2-5 min) and slow cooling to 37 °C for proper RNA folding. Same amount of PF4 and RNA were mixed with sodium chloride (final concentration of NaCl 600 mM). Reaction took place at 37 °C for 30 min. Sample was dropped on freshly cleaved mica, adsorption took place within 2 min and was followed by washing of mica surface with millipore for 30 s and drying in a laminar flow for ca. 15 min. Fresh or maximal one month old samples were imaged as described in section AFM imaging.

AFM imaging

AFM imaging was done using Leica DMI 6000B microscope, laser beam, standard cantilevers model TESP and nanoscope software 7.31, applying standard tapping mode in air in advanced mode.

Results:

We performed two separate selection processes to obtain Aptamers with distinct binding properties (Figure 1). The first procedure was carried out under the focus of selecting Aptamers that are prone to form large, multimolecular complexes similar to those formed with heparin and PF4. The interactions between heparin and PF4 are predominantly electrostatic. Thus, the selected Aptamers sharing the same features as heparin and forming large complexes with PF4 are likely to bind PF4 via electrostatic interactions as well.

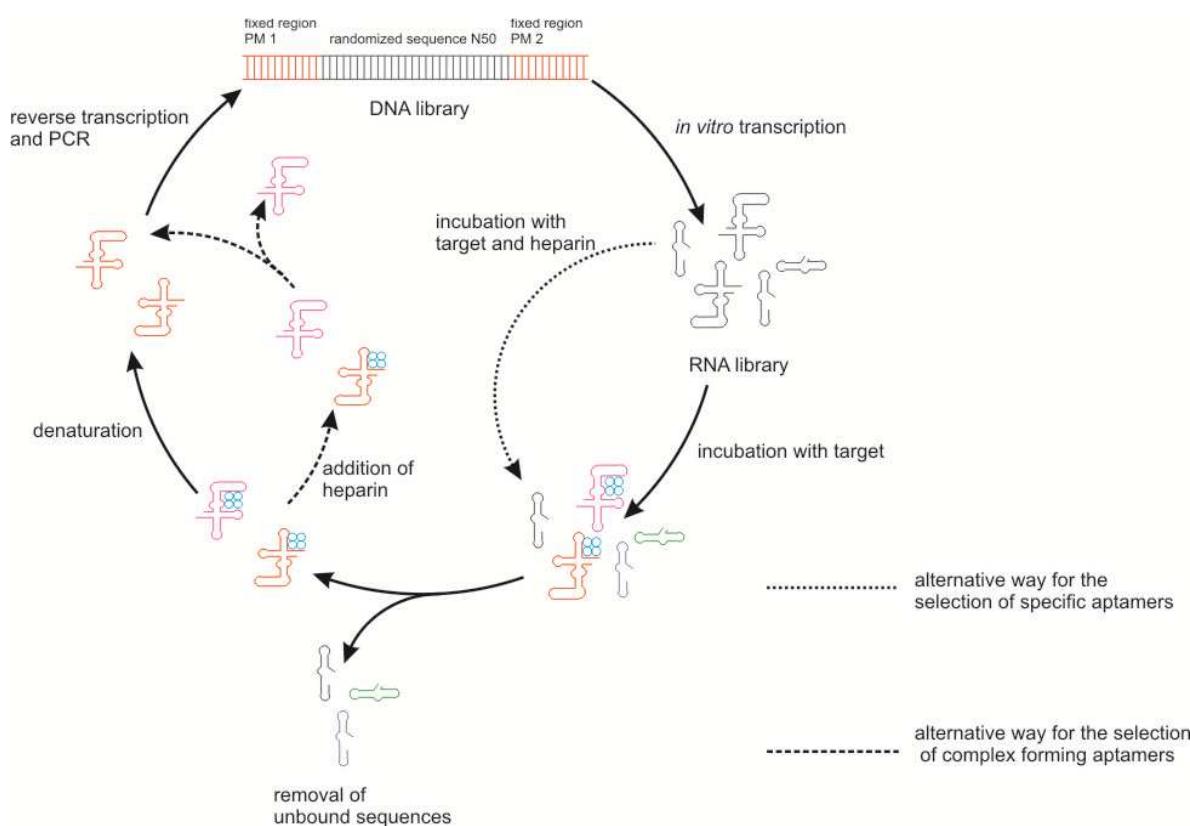


Figure 1: Selection strategies for the selection of both complex forming and specific RNA Aptamers for binding to PF4.

Selection of complex forming Aptamers

The selection was performed in 25 mM Tris-acetate buffer supplemented with 5 mM MgCl₂ using a RNA library comprising a sequence of 50 randomized nucleotides. Prior to the first round of selection, the library was incubated with different amounts of PF4 to determine the optimal starting conditions in terms of the RNA-PF4 ratio. To start the selection, the initial RNA library was incubated with PF4. After each round, complexes were separated from free RNA and protein by centrifugation using filters with an exclusion size of 50 kDa. To obtain sequences that form the most stable complexes with PF4, the amount of RNA and the reaction temperature was varied in later rounds of the process. Additionally, heparin was added as competitive inhibitor of RNA-PF4 interactions. At the end of each round, separation of RNA and PF4 was achieved by adding 1M NaCl solution. Since NaCl is known to disrupt interactions between PF4 and polyanions (Visentin, Ford, Scott & Aster 1994) it was added to the selection buffer from round nine on to apply more selection pressure. After 13 rounds the final pool was cloned and sequenced.

Sequence analysis delivered 26 different sequences. An alignment of the sequences revealed only minor families along with orphan sequences. Therefore, we shifted our focus away from the primary structure towards similarities in the secondary structure. Folding analysis (using RNAstructure 4) revealed three major secondary structure motifs. The majority of the selected Aptamers folds into three- and four way junctions, along with a minority of rod-shaped structures.

Next, we analyzed the obtained Aptamers for their potential to form complexes with PF4 using gel shift assays (Figure 2). Complexes formed between PF4 and the Aptamers did not enter the gel and remained stationary in the gel pockets. Thus, RNA present in gel pockets indicates complex formation with PF4. In order to determine best binders, the Aptamers were incubated with PF4 at different buffer compositions. Selection buffer supplemented with 600 mM MgCl₂ served as positive control (Figure 2A), whereas selection buffer containing 1 M NaCl was used as negative control (Figure 2C). Additionally, the Aptamers were incubated with PF4 in the presence of 800 mM NaCl (Figure 2B). Under these conditions only five Aptamers were still capable of binding PF4 (Figure 2B, for Aptamer sequences see table 1).

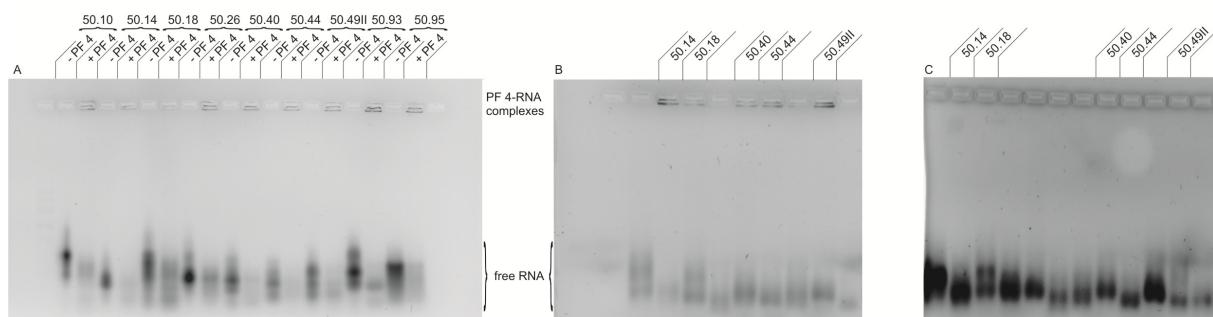


Figure 2: Complex formation of the selected Aptamers with PF4 at 600 mM NaCl (A), 800 mM NaCl (B) and 1 M NaCl (C).

Binding of these five Aptamers was further analyzed by photon correlation spectroscopy (PCS; Figure 3). The five Aptamers, 50.14, 50.18, 50.40, 50.44 and 50.49II, were incubated with PF4 at NaCl concentrations varying from 0.6 to 1 M. An increase of the NaCl concentration is supposed to disrupt the formed complexes. Hence, more stable complexes should disrupt only at higher NaCl concentrations. In the PCS experiment, particles with a diameter of 5-10 nm were considered as PF4 tetramers (Greinacher *et al.* 2006). The results do not indicate major differences between individual Aptamers. However, in all cases, the amount of free PF4 is less than 50%, indicating that all five Aptamers form large, stable complexes with PF4, even at rather high NaCl concentrations. Solely Aptamer 50.18 showed least variation of free PF4 at high NaCl concentrations.

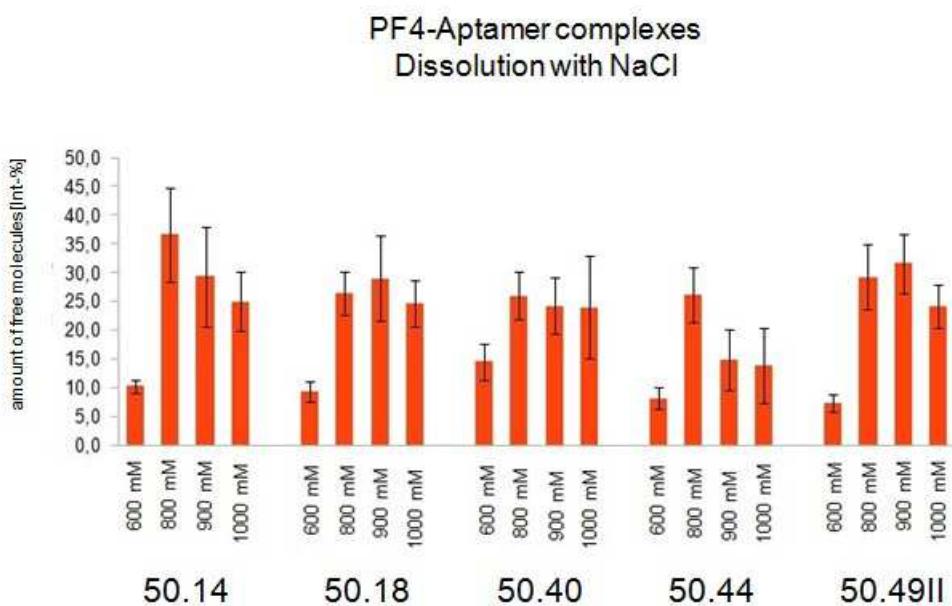


Figure 3: PCS analysis of Aptamer-PF4 complexes. Shown is the amount of free PF4 at varying NaCl concentrations.

Table 1: Sequences of the selected Aptamers which form the most stable complexes

Name	sequence (5' → 3')	Length
50.14	GGGAGAGCCAUCUACCUGACUUACCAAAUAGGUGAAUGGUG UGACAGACAAACUGCUUACCUUGAGUUUGUCAGUCGUCAGG AUCCGUG	87
50.18	GGGAGAGCCAUCUACCUGACUCAGGUUUCUCGACCUGUACCC	88

	UUCGCGGUACAUUCUGUCGAACUGGUCGUCAGUCGUCAG GAUCCGUG	
50.40	GGGAGAGCCAUCUGACUUCCCGGCUGGUCCGGAGUACG AACUACACCUGGCCAUGAAUUGUCGGCGGUCAGUCGUCAG GAUCCGUG	88
50.44	GGGAGAGCCAUCUGACUCUUUGAUUGGGUGUUGUAGC CCCAAAGCGGGCCUUGGGGUACUGAGUGUCAGUCGUCAG GAUCCGUG	88
50.49II	GGGAGAGCCAUCUGACUUCCCGGAGAACCGACGUCUCU CUGAAAAACGUUCAAAGAGAGCGUCCGUUAGUCAGUCGUCAG GAUCCGUG	88

Selection of specific Aptamers

In a second procedure we aimed for selecting Aptamers that bind PF4 specifically using the same library as in the first selection but changing the buffer system. Since specific Aptamers might potentially be used in cell cultures, we chose a buffer system that contained the most important cations in physiological concentrations: 25 mM Tris-HCl supplemented with 150 mM NaCl; 5 mM KCl; 2.5 mM CaCl₂ and 1 mM MgCl₂. After incubation of the library with PF4, unbound sequences were removed by filtering the reaction volume through a nitrocellulose membrane. The selection was started with the pool obtained after the third round of the previously described selection, which had already been enriched for sequences that bind PF4, in order to circumvent the otherwise necessary pre-selection. During the process, the stringency was increased by the addition of heparin, by varying the PF4-RNA ratio and by more vigorous washing. After eight rounds of selection, isolation and amplification the final pool was cloned and sequenced.

Out of a total of 96 clones, 60 different sequences were obtained. For further investigation we chose the five sequences with the highest copy number in the final pool. First, we analyzed the binding characteristics for the five Aptamers (see table 2 for sequences) by electrophoretic mobility shift assays in native polyacrylamid gels (Figure 4).

As shown in Figure 4, bands in the upper parts of the gels are visible upon staining with SYPRO Orange (protein sensitive) and ethidium bromide (nucleic acid sensitive), indicating that those bands representing both, protein and RNA, and thus the PF4-Aptamer complexes. Furthermore, there seem to be Aptamer-PF4 complexes of variable global structure and/or stoichiometry with two dominating species, represented by two pronounced bands in the upper part of the gels. In general, a different behavior of the specific Aptamers compared to the complex

forming Aptamers described above is observed. After binding to PF4, the specific Aptamers were still capable of moving into the gel, whereas in case of the complex forming Aptamers, bound RNA was only detectable in the gel pockets, indicating that the specific Aptamers do not support the formation of large, multimolecular complexes with PF4.

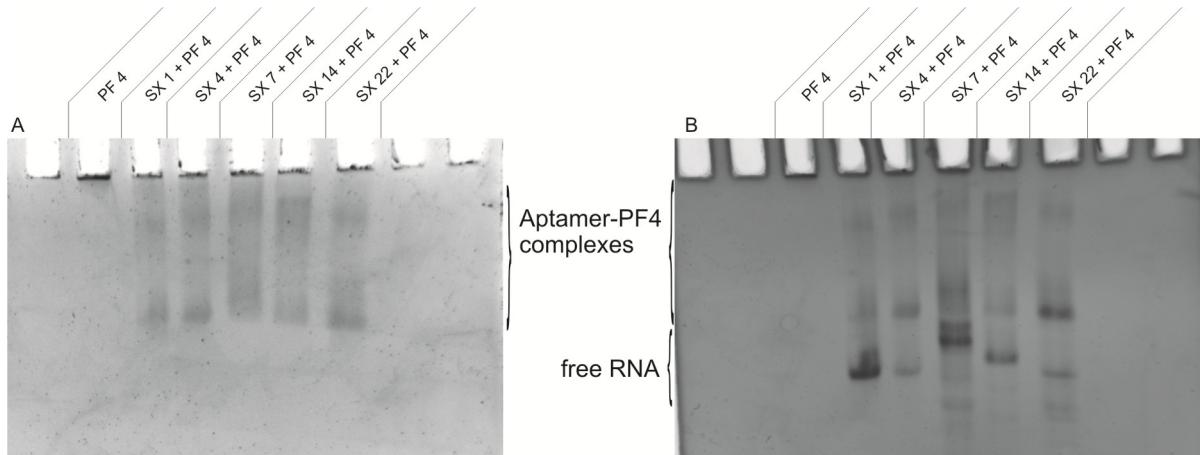


Figure 4: Binding of specific Aptamers to PF4. The gels were stained with SYPRO Orange (A) or ethidium bromide (B).

Table 2: Sequences of specific PF4 Aptamers

name	sequence (5' → 3')	Length
SX 1	GGGAGAGCCAUCUACCUGACUGGCCUUUCUACUGUUUUCCAGU UACGCCGUACGGCGCAGAAGGUAGCAGUCAGUCGUCAGG AUCCGUG	88
SX 4	GGGAGAGCCAUCUACCUGACAUUGUCUCCCGAUUGAGGGAGGC GAGUCAUCGAUCGAUCUCGGUCAGUCGUCAGG AUCCGUG	87
SX 7	GGGAGAGCCAUCUACCUGACAAGGGGUUCCCUUGCUUGUGCGA UCAACUUCAUAGUCUCGCACACGCGUUGUCAGUCGUCAGGA UCCGUG	88
SX 14	GGGAGAGCCAUCUACCUGACAUCUCGUUUUUGAGAACAGGUC CCGGGACUGAGGAUGCAAGGUCCUAACAGUCAGUCGUCAGGA UCCGUG	88
SX 22	GGGAGAGCCAUCUACCUGACAGGAAUCUCCCCAAAAACGGACUG GAUGAACCCGACCGGACAUCGAGCAGUCAGUCGUCAGGAU CCGUG	88

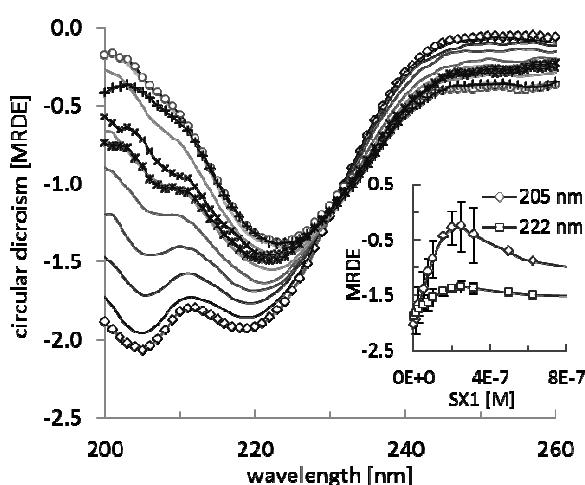
Next, we looked at possible conformational changes of PF4 upon Aptamer binding by CD-spectroscopy. In addition, concentration-dependent measurements were conducted in order to determine the binding affinity. We also used surface plasmon resonance to determine binding parameters. However, these measurements were

not successful due to problematic properties of PF4: i) PF4 tends to self-aggregate and showed significant binding to various surfaces, which could not be fully prevented. Therefore, immobilization of the Aptamers and utilizing PF4 as analyte was impossible. ii) Immobilizing PF4 and using the Aptamers as analytes also failed. The problem was the generation of a stable surface. PF4 having a homotetrameric structure showed considerable dissociation after each washing step of the surface. We tried a number of different immobilization strategies for both PF4 and the Aptamers (by biotin-streptavidine interactions and aliphatic amine-succinimidyl ester coupling). However, no set up delivered convincing and reproducible data.

Therefore, CD-spectra measured at different PF4/Aptamer molar ratios (constant concentration of PF4, subsequent increased Aptamer concentration) were inspected for characteristic changes in dependence with the Aptamer concentration (Figure 5). For initial experiments Aptamers SX1 and SX4 were used. For both of these Aptamers pronounced changes in the CD-spectra towards more positive ellipticity values are observed with increasing Aptamer concentrations (Figure 5A). This trend continues until a certain PF4/Aptamer ratio is reached (2.5×10^{-7} M for both SX1 and SX4), and reverses by further increasing the Aptamer concentration (Figure 5B). This becomes most obvious at 205 and 222 nm (inserts in Figure 5). These two bands are usually attributed to the presence of α -helical secondary structures. The absolute values at 205 and 222 nm are proportional to the α -helix content of the protein (Toumadje, Alcorn *et al.* 1992; Sreerama and Woody 2004). The insets of Figures 5A and 5B show lowering in the absolute values in these two bands towards the critical ratio, indicating on qualitative level a reduction of the α -helical secondary structure content.

Preliminary figures

A



B

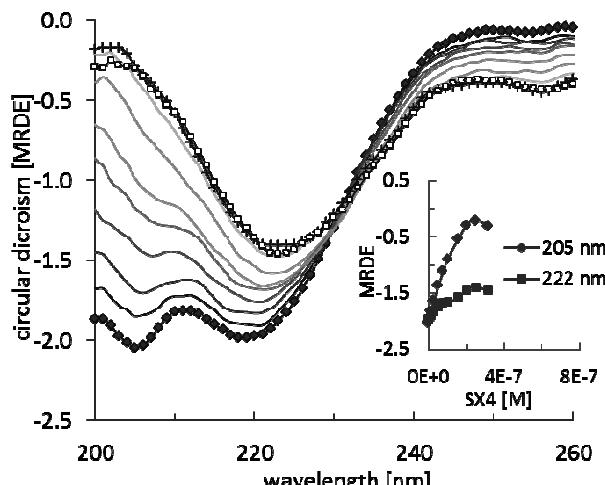


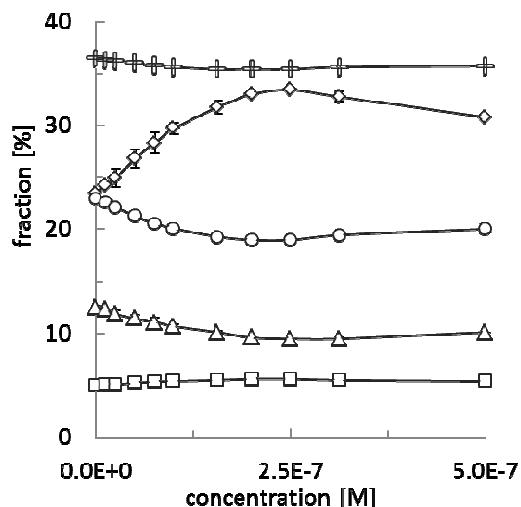
Figure 5: CD spectra of PF4/RNA complexes formed at different RNA concentrations: A – the concentration range of SX1 was $0 - 1 \times 10^{-6}$ M whereas in B – the SX4 concentration was stepwise increased from $0 - 3.1 \times 10^{-7}$ M. The insets show

the development of ellipticity at 205 and 222 nm, the two bands that are typically attributed to the presence of α -helical structures. —○— native PF4, —○— PF4 + RNA (2.5×10^{-7} M = point of maximum structural changes), —+— PF4 + RNA (3.1×10^{-7} M), —*— PF4 + RNA (5.0×10^{-7} M), —**— PF4 + RNA (6.3×10^{-7} M).

Deconvolution of the CD-spectra allowed the quantification of PF4 structural changes in detail (Figure 6). By increasing the RNA concentration, a pronounced decrease in α -helical content occurs and the maximum changes are observed for a RNA concentration of 2.5×10^{-7} M. In parallel to the decrease in the α -helical content, deconvolution suggests a small decrease of β -turn structures. This decrease is balanced by a pronounced increase in the antiparallel β -sheets content. Other secondary structures undergo only insignificant changes. According to the obtained curves we were able to determine the K_D -value for Aptamer SX1 with 78 nM.

Preliminary figures

A



B

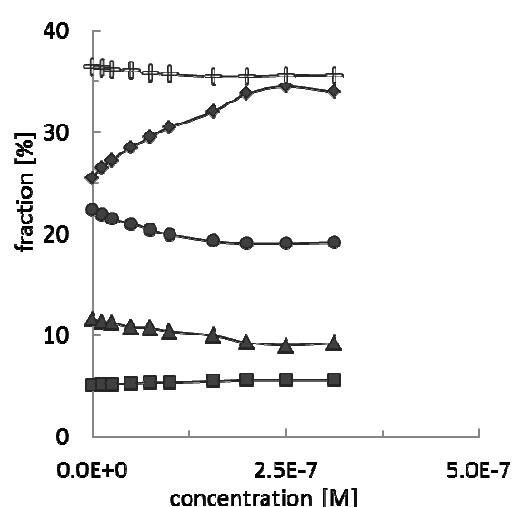


Figure 6: Deconvolution of the CD spectra provides the secondary structure content of the protein at each PF4/RNA molar ratio. A – for PF4/SX1, B – for PF4/SX4 (—○— α -helix, —●— $\uparrow\downarrow\beta$ -sheet, —■— $\uparrow\uparrow\beta$ -sheet, —◆— β -turn, —+— random coil).

For single-molecule illustration with AFM of two different specific complexes with PF4 RNAs 50.40 and SX4 were chosen.

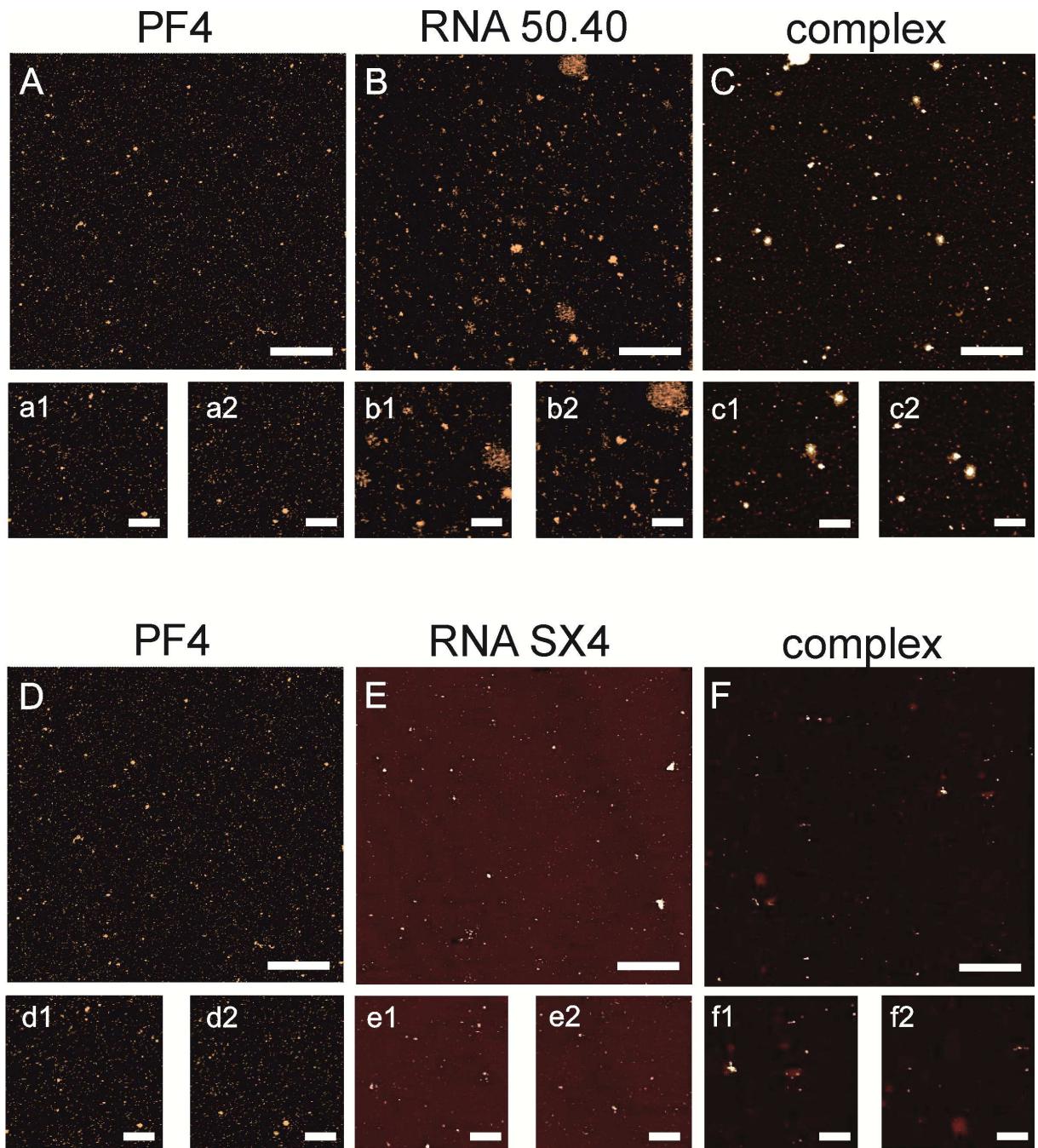


Figure 7: AFM data

AFM tapping mode in air illustration of PF4 (A, D), aptamer 50.40 (B), SX4 (E), complexes composed of PF4 and aptamer 50.40 (C), SX4 (F). Scale bars in images A-F represent 2 μm , scale bars a1-f2 represent 600 nm.

For complexes including SX4 small features are found, presumably composed of one to three PF4 tetramers and surrounding PF4, with a length of maximal 250 nm. Complexes including RNA 50.40 displayed a wide range of different shapes sizes, such as linear to globular or ellipsoidal in a range of few nanometers to 1.1 μm (fig. 8).

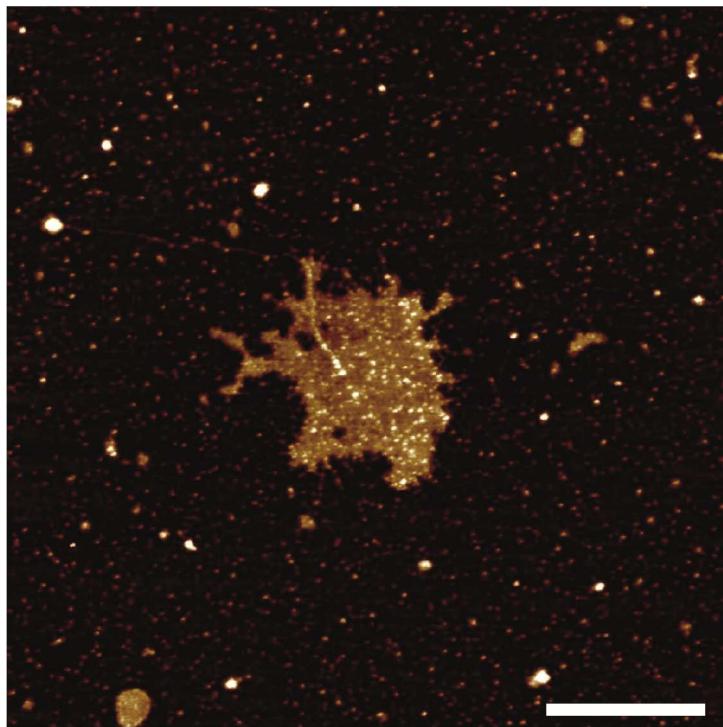


Figure 8: AFM tapping mode illustration of PF4-50.40 complexes

Scale bar 1 μm ; several complexes of different sizes and shapes.

Discussion

We have performed two parallel selections for Aptamers binding to PF4, using identical libraries, though with a different focus on the outcome. The first selection yielded Aptamers that are prone to form large complexes with PF4. Although these sequences do not share apparent nucleotide similarities, they all fold predominantly into three- and four way junctions. The lack of sequence families could be explained by the fact that the interactions between these Aptamers and PF4 are predominantly of electrostatic nature. Therefore the sequence is probably less important, whereas the three-dimensional structure and thus, the positioning of the negatively charged phosphate groups plays a superior role in binding to PF4 and complex formation. Since we were looking for Aptamers capable of forming multimolecular complexes, the outcome that most Aptamers fold into three- and four way junctions is not surprising. Investigations with branched glucan sulfates showed that these branched carbohydrates support complex formation with PF4 at lower concentrations compared to heparin (Greinacher *et al.* 1995). In terms of the shape, three- and four way junctions might be more comparable to branched carbohydrates, whereas less structured rod shaped species display more similarities with heparin. Furthermore, the presence of multiple helices could facilitate the formation of large PF4-RNA networks, as we suggest in the model of Aptamer-PF4 interaction shown in Figure 7. Apparently, the more helices an Aptamer offers, the more PF4 molecules can interact with it. In addition, folding of the Aptamers might result in tightening the RNA-PF4

interaction, which would make the complexes more compact and highly resistant to dissolution in high salt buffers (up to 900 mM NaCl).

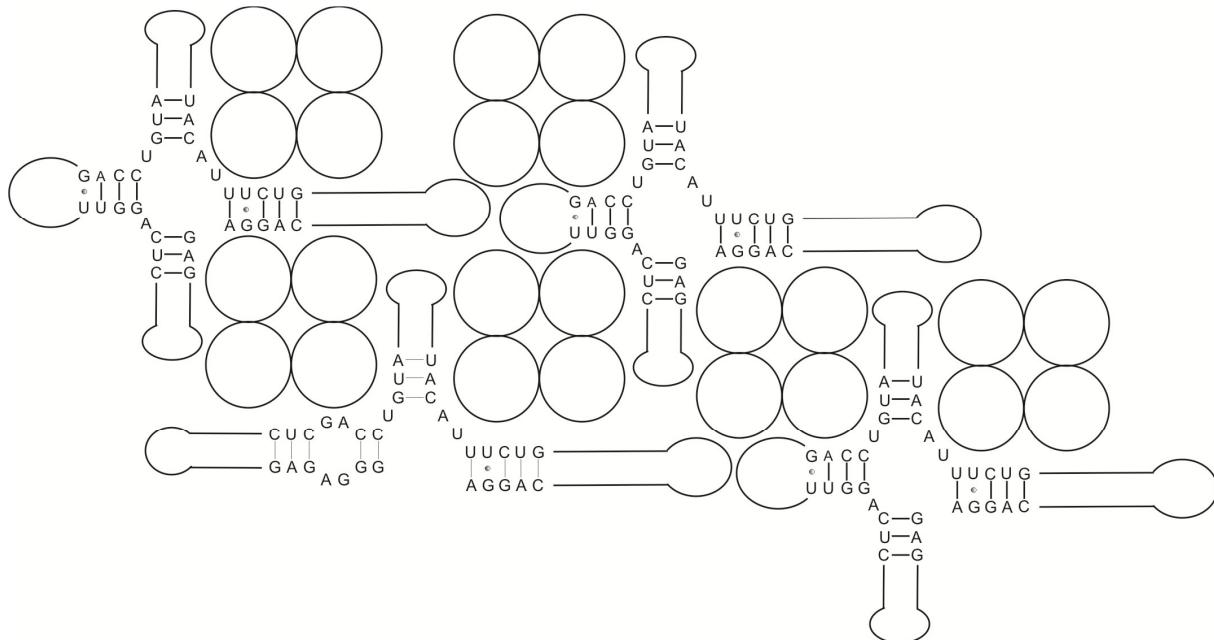


Figure 7: Model of complex formation between PF4 and the selected Aptamers.

The selected Aptamers showed only minor differences in complex stability. Only five Aptamers, however, were able to form complexes beyond a concentration of 600 mM NaCl, which was used in the final selection round. Among these five Aptamers, Aptamer 50.18 showed least variation of free PF4 at higher NaCl concentrations as detected by photon correlation spectroscopy, indicating that it forms most stable complexes. The secondary structure simulation for this Aptamer revealed three- and four way junction structures, which explains its capability of forming large, tight PF4-Aptamer complexes that are stable up to 900 mM NaCl.

The second selection under the premise of finding specific Aptamers for PF4 resulted in Aptamers with distinct binding properties as the complex forming Aptamers. In gel shift assays, we showed that these Aptamers do not form large complexes with PF4, nevertheless displaying strong binding. Their different binding characteristics can be considered as a result of the selection process, in which they were evolved. Using nitrocellulose filter binding prevents sequences from being enriched that strictly interact electrostatically. Since the filter surface is negatively charged PF4 will orientate in a way that its positively charged clusters interact predominantly with the filter and become not available for the RNA. Thus Aptamers selected here should bind PF4 first and foremost via specific interactions and not by electrostatic attraction.

Using CD-spectroscopy we were able to identify structural changes of PF4 upon binding of Aptamers SX1 and SX4. The results indicate that PF4 undergoes

structural changes that include a decrease in the α -helical content and β -turns, accompanied by a significant increase in the antiparallel β -sheet content. Recently, we observed these features also for heparin as ligand instead of RNA (Jaax *et al.* 2013). The structural changes emerge presumably due to the binding pattern. Upon binding to the Aptamers or heparin, respectively, PF4 tetramers are brought in close vicinity. Tightening up the Aptamers might lead to steric hindrance and electrostatic repulsion which in turn leads to conformational changes of PF4. With the help of the CD-data the K_D -value for binding of Aptamer SX1 to PF4 was determined with 78 nM, which is well in the range of affinities measured for binding of a variety of Aptamers to their respective ligands (Nimjee; Rusconi & Sullenger 2005).

Conclusion:

As a result of this work, complex-forming and specific Aptamers to PF4 were developed. This will allow further addressing the question if particular Aptamers may cause an immune response when applied as drugs, analogous to what is known as heparin-induced thrombocytopenia (HIT). Further studies will include immunological analysis of the selected complex-forming Aptamers, in particular investigating if those PF4-Aptamer complexes induce the same immune response as heparin. This has significant impact on the field of RNA therapeutics, suggesting that three- and four way junctions in therapeutic Aptamers need to be avoided.

The selected Aptamers that bind specifically to PF4, are now available for further studies on the function of PF4, and in addition, might also be used for medical applications. As competitive binders to heparin, those Aptamers might be used as co-drugs to prevent complex formation between heparin and PF4 and the subsequent immune response. This, in turn, would allow avoiding serious side effects when administration of heparin is indicated.

References:

- Deuel, T F; Senior, R M; Chang, D; Griffin, G L; Heinrikson, R L; Kaiser E T *Proc. Natl. Acad. Sci.* **1981**, 78(7), 4584-4587
- Deuel; T F; Keim, P S; Farmer, M; Heinrikson, R L. *PNAS* **1977**, 74, 2256-2258
- Greinacher, A; Pötzsch, B; Amiral, J; Dummel, V; Eichner, A; Mueller-Eckhardt, C *Thrombosis and Haemostasis* **1994**, 71(2), 247-251
- Greinacher, A; Alban, S; Dummel, V; Franz, G; Mueller-Eckhardt, C *Thrombosis and Haemostasis* **1995**, 74(3), 886-892

Greinacher, A; Gopindhan, M; Günther JU; Omer-Adam, M.A.; Strobel, U; Warkentin, T.E.; Papastavrou, G; Weitschies, W; Helm, C.A. *Arterioscler Thromb Vasc Biol.* **2006**, 26, 2386-2393

Jaax, M E; Krauel, K; Marschall, T; Brandt, S; Gansler, J; Fürll, B; Appel, B; Fischer, S; Block, S; Helm C A; Müller, S; Preissner K T; Greinacher A *Blood* **2013**

Kaplan, K. L.; Broekman, M. J.; Chernoff, A.; Lesznik, G. R.; Drillings, M. *Blood* **1979**, 53, 604-618.

Krauel, K; Weber, C; Brandt, S; Zähringer, U; Mamat, U; Greinacher, A; Hammerschmidt, S *Blood* **2012**

Nimjee, S.M.; Rusconi, C.P.; Sullenger, B.A. *Annu. Rev. Med.* **2005**, 56, 555-583

Rauova, L.; Poncz, M.; McKenzie, S. E.; Reilly, M. P.; Arepally, G.; Weisel, J. W.; Nagaswami, C.; Cines, D. B.; Sachais, B. S. *Blood* **2005**, 105, 131-138

Rucinski, B; Niewiarowski, S; James, P; Walz D A; Budzynski, A Z. *Blood* **1979**, 53, 47-62

Senior, R M; Griffin, G L; Huang, J S; Walz, D A; Deuel, T F *The Journal of Cell Biology* **1983**, 96, 382-385

Sreerama, N. and R. W. Woody. *Numerical Computer Methods, Pt D* **2004** 383, 318-351.

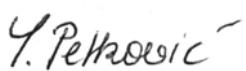
Toumadje, A., S. W. Alcorn, et al. *Faseb Journal* **1992**, 6(1), A350-A350.

Warkentin, T. E.; Greinacher, A. *Chest* **2004**, 126, 311s-337s

Anteile der Autoren an den Publikationen:

- 1) Müller, S., Appel, B., Krellenberg, T. and Petkovic, S. (2012) The many faces of the hairpin ribozyme: structural and functional variants of a small catalytic RNA. *IUBMB Life*, **64**, 36-47
S.P. betrieb die Literaturrecherche, fertigte die Abbildungen 5, 9 und 10 an, diskutierte und kommentierte das Manuscript. B.A. und T.K. erstellten die übrigen Abbildungen. S.M. verfasste das Manuscript.
- 2) Petkovic, S. and Müller, S. (2013) RNA self-processing: Formation of cyclic species and concatemers from a small engineered RNA. *FEBS Lett*, **587**, 2435-2440
S.P. synthetisierte Klenow-DNA, synthetisierte die RNA, führte die Ribozym-Reaktionen und alle Enzymreaktionen durch, schrieb die Produkt-RNAs in cDNAs um, vervielfältigte sie ligierte, klonierte DNA und analysierte und interpretierte die entsprechenden Sequenzdaten. S.P. führte die exoribonukleolytische Hydrolyse und die 2D-Analysen durch. S.P. fertigte einen Argumentationsfaden und einen ersten Entwurf für das Manuscript an. S.P. und S.M. diskutierten das Manuscript. S.M. erstellte die finale Version des Manuscripts.
- 3) Petkovic, S.#, Badelt, S. # (shared first authorship), Block, S., Flamm C., Hofacker, I., Delcea, M., Müller, S., (2014), Sequence-controlled RNA self-processing: computational design, biochemical analysis and visualization by AFM, *JACS* Manuscript eingereicht am 14. März 2014, Manuscript ID: ja-2014-02618m
S.P. führte alle biochemischen experimentellen Arbeiten, angefangen bei der DNA-Synthese, zur RNA-Synthese, Ribozym-Reaktionen, Synthese und Bereitstellung der RNA für die AFM-Messungen, exonukleolytischen Abbau, bis zur 2D-Analyse aller Test-Systeme und Referenzsysteme durch. S.P. verfasste einen Argumentationsfaden für das Manuscript und erstellte das erste Manuscript mit S.M. S.Ba., C.F. und I.H. führten die bioinformatischen Arbeiten durch. S.BI. führte die AFM-Messungen durch. Alle Autoren diskutierten das Manuscript und S.P. und S.M. verfassten die finale Version.
- 4) Petkovic, S. and Müller, S. (2014), Generation of circular RNA *in vivo* and *in vitro*, *Nucleic Acids Research*, eingereicht am 26. März 2014, Manuscript ID: NAR-00828-2014
S.P. hat den Entwurf, die Literaturrecherche, alle Abbildungen und alle Manuscript-versionen erstellt. S.M. hat das Manuscript überarbeitet und mit S.P. diskutiert.
- 5) Marschall, T., Brandt, S., Kreimann, M., Petkovic, S., Frommer, J., Greinacher, A., Delcea, M., Müller, S. (2014), Evolution and Characterization of Aptamers targeting Platelet Factor 4, *Nucleic acid research*, Manuscript in Vorbereitung
S.P. bereitete die RNA-, PF4-, und RNA-PF4-Proben für die AFM vor. Dazu gehörte die Komplexbildung der RNA mit PF4. S.P. erstellte die AFM-Abbildungen und schrieb die entsprechenden Teile des Manuscripts. T.M. führte die SELEX, die Sequenzierungen und *Gelshift-Assays* durch und verfasste das initiale Manuscript ohne den AFM-Teil. S.B. führte die CD-Experimente durch. M.K. führte die PCS-Messungen durch. J.F. stellte die RNAs 50.40 und SX4 in großem Maßstab her. A.G., M.D. und S.M. haben das Manuscript überarbeitet. Alle Autoren diskutierten das Manuscript.

Betreuerin



Sonja Petkovic



Prof. Sabine Müller

Publikationsliste

Forschungsartikel und Reviews

Müller, S., Appel, B., Krellenberg, T. and Petkovic, S. (2012) The many faces of the hairpin ribozyme: structural and functional variants of a small catalytic RNA. *IUBMB Life*, **64**, 36-47

Petkovic, S. and Müller, S. (2013) RNA self-processing: Formation of cyclic species and concatemers from a small engineered RNA. *FEBS Lett*, **587**, 2435-2440.

Petkovic, S.#, Badelt, S.# (shared first authorship), Block, S., Flamm C., Hofacker, I., Delcea, M., Müller, S., (2014), Sequence-controlled RNA self-processing: computational design, biochemical analysis and visualization by AFM, *JACS* manuscript submitted: 14th March 2014, Manuscript ID: ja-2014-02618m

Petkovic, S. and Müller, S. (2014), Generation of circular RNA *in vivo* and *in vitro*, *Nucleic Acids Research*, submitted 26-Mar-2014: Manuscript ID: NAR-00828-2014

Marschall, T., Brandt, S., Kreimann, M., Petkovic, S., Frommer, J., Greinacher, A., Delcea, M., Müller, S. (2014), Evolution and Characterization of Aptamers targeting Platelet Factor 4, *Nucleic acid research*, manuscript in preparation

Vorträge

Sonja Petkovic: 2012, Self-processing ciRNAs: Formerly RNA genomes, nowadays therapeutically tools?! GMB-Meeting RNA biochemistry and Workshop RNA trafficking, Bonn, 04.-07. 10. 2012

Sonja Petkovic: 2013, Self-processing RNA: Computationally designed – experimentally proven, 28. TBI Winterseminar in Bled, Slowenien & 7th Annual Meeting of Bompfünnewerer Consortium February 2013

Sonja Petkovic: 2013, Analyses of self-processing RNA, Erstes Institutsseminar, Institute of Biochemistry, Greifswald

Poster

Sonja Petkovic, Stephan Block, Mihaela Delcea and Sabine Müller: 2013, Identification of circRNAs in the product mixture of a self-processing RNA; IV Nukleinsäurechemie-Treffen 19./20.-09. 2013

Eidesstattliche Erklärung

Hiermit erkläre ich, dass diese Arbeit bisher von mir weder an der Mathematisch-Naturwissenschaftlichen Fakultät der Ernst-Moritz-Arndt-Universität Greifswald noch einer anderen wissenschaftlichen Einrichtung zum Zwecke der Promotion eingereicht wurde.

Ferner erkläre ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die darin angegebenen Hilfsmittel und Hilfen benutzt und keine Textabschnitte eines Dritten ohne Kennzeichnung übernommen habe.



S. Petkovic

Danksagung

Zuallererst möchte ich meiner „Doktormutter“ Prof. Sabine Müller danken. Ich möchte mich bedanken, dafür dass ich aus vielen möglichen Projekten ein für mich besonders interessantes auswählen durfte, dass ich für alle Fragen rund um Promotion stets Zeit, Rat und Tat erhalten habe sowie für das angenehme Vertrauensverhältnis, dass sich während meiner Tätigkeit entwickelt hat.

Zudem möchte ich dem gesamten Arbeitskreis Prof. Müller für jegliche Art von Unterstützung danken, von der angesetzten Gellösung bis zur spendierten Schokoladentorte.

Danken möchte ich auch den Kooperationspartnern aus Österreich Prof. Hofacker, Prof. Flamm für die konstruktiven Gespräche in Wien, Berlin und Bled, Dr. Höner zu Siederissen für die freundliche Aufnahme und in Wien und die Benennung von Arbeitsgruppen zur Rasterkraftmikroskopie. Weiter gilt mein Dank dem Promotionsstudenten Stefan Badelt für das Entwerfen der in dieser Arbeit getesteten CRZ-2-Varianten, für vielfachen Austausch zu seiner bioinformatischen Vorgehensweise und Diskussion der Ergebnisse und für die Einladung nach Slowenien zum Vortrag.

Danken möchte ich auch Dr. Mihaela Delcea und Dr. Stephan Block als Kooperationspartner zur Untersuchung von RNAs mit AFM. Zudem möchte ich beiden Danken für eine eigene Einführung zu AFM, was mir ermöglichte selbst RNA-AFM-Bilder im Arbeitskreis von Dr. Delcea zu erstellen. Dafür gilt auch insb. mein Dank dem Promotionsstudenten Martin Kreimann, der mir den Umgang mit dem AFM, Probenvorbereitung, Datenillustration und Auswertung beigebracht hat.

Aus dem Arbeitskreis Prof. Müller möchte ich einige Kollegen hervorheben und ihnen danken, wie Anne Strahl, die mir einen methodischen Einstieg mit Ribozymen ermöglicht hat und immer für Fragen offen war, Kristina Salomon für ihre Hilfe mit den Sequenzierungen, Vektor NTI und PDE-Größenbestimmung, Danilo Springstubbe als MacGyver des Arbeitskreises, sowie Maria Jenckel, Madlen Wolf, Marion Groth, Rita Schröder und Simone Turski, die mir die Zeit entweder durch gemeinsame Kakaomomente oder durch ein gutes Wort hier und da haben kurzweilig erscheinen lassen.

Ich möchte auch meinen Praktikanten Eugen Makuha, Birte Fredrich, Madlen Wolff und Carina Spormann danken, denn ich konnte immer als Lehrende auch etwas lernen.

Zu guter Letzt möchte ich meinen Freunden Małgorzata Gac und Silke Wiesner, meiner Schwester Fabienne Kracke und meinem Papa Mirko Petković für Ihre Unterstützung auf meinem Weg danken.

Danke lieber Papa, du bist mein Vorbild.