RNA Chemistry

DOI: 10.1002/anie.201205427

How a Chemist Looks at RNA

*Thomas R. Cech**

chemical genetics · kinetics · RNA · thermodynamics

I was passing through customs some time ago, returning from London. The immigration officer, seemingly unsatisfied with my self-identification as "professor," inquired about my research area. "Molecular biology," I answered; then added "ribonucleic acid," certain this would convince him that I was harmless.

At other times I identify myself as a biochemist, but almost never as just-plain chemist. This may seem a bit odd, given that all of my formal degrees (B.A. Grinnell College, Ph.D. UC Berkeley) are in Chemistry, my first (and only) faculty appointment has been in a Department of Chemistry, and I teach General Chemistry to undergraduates. The opportunity to write something to help celebrate 125 years of *Angewandte Chemie* has prompted me to organize some thoughts about the importance of chemistry in my own research and in RNA research more generally. But first, in case there are any immigration officers reading this piece, what's so interesting about RNA?

A Brief History of RNA Research

In the 1960s and for most of the 1970s, students knew that there were three kinds of RNA: messenger RNA (mRNA), which carries the genetic code from DNA to protein, and ribosomal RNA (rRNA) and transfer RNA (tRNA), which are involved in the read-out of the mRNA into protein. As more noncoding RNAs were discovered, it became clear that the functionality of RNA is much broader than contributing to protein synthesis. First, RNA can be a biological catalyst. An intron RNA from *Tetrahymena* self-catalyzes its own site-specific cleavage and splicing reactions[1] and in the process specifically recognizes a small-molecule ligand (guanosine) and uses it as a nucleophile.[2] RNase P RNA acts as a multiple-turnover catalyst, processing tRNA precursors by site-specific hydrolysis.[3] Small nuclear RNPs (RNA–protein complexes[4]) recognize specific splice-site sequences on precursors to mRNA and then participate directly in the splicing reaction.[5] At the same time, it became increasingly clear that rRNA must have an active role in protein synthesis,[6] which was later confirmed in atomic detail by X-ray crystallography.[7] Evolution-in-a-test-tube experiments showed that the potential versatility of RNA far exceeded the examples found in nature thus far,[8] adding credibility to the idea of a primordial "RNA World" where RNA served as both genetic material and as catalyst.

The next rush of RNA discoveries concerned RNA as a regulator of gene expression. It had previously been thought that most regulation was conducted by proteins that either repressed or activated gene transcription (the DNA→RNA step) or regulated mRNA stability or translation (the RNA→protein step). But it turned out that especially in Gram positive bacteria, many mRNAs have built-in "riboswitches" that bind small-molecule metabolites and switch gene expression on and off.[9] This finding was presaged by the attenuation mechanism, which also involves RNA conformational switches.[10] In eukaryotes, microRNAs down-regulate whole groups of mRNAs at the same time by base-pairing to sequences in the 3′ untranslated region (the region that follows the string of codons).[11] In female mammals, the condensation and functional inactivation of one of the two X chromosomes is mediated by long noncoding RNAs,[12] and more localized examples of long noncoding RNA (lncRNA) repression of gene expression have been found by the thousands.[13]

Such has been the explosive nature of RNA research within the last two decades that the above summary could easily be extended to include numerous other examples, such as the telomerase chromosome end-replicating RNP,[14] small interfering (si) RNAs, IRES RNAs that allow translation in the absence of a 5′-cap on the mRNA, scan RNAs that mediate DNA elimination during macronuclear development in ciliates, CRISPR RNAs that mediate bacterial defense against invading nucleic acids, and RNAs that mediate homologous chromosome association during meiosis.

The Importance of Being Chemical

Chemistry makes a special impact on RNA research in two rather different ways. 1) A chemist brings a useful attitude to the interrogation of RNA mechanisms, an attitude that is quantitative, steeped in appreciation of kinetics and thermodynamics, and based on molecular structure. 2) A chemist has the ability to synthesize RNA molecules with

[*] Prof. T. R. Cech
Director, University of Colorado BioFrontiers Institute
Investigator, Howard Hughes Medical Inst.
Department of Chemistry and Biochemistry
University of Colorado
Boulder, CO 80309-0596 (USA)
E-mail: Thomas.cech@colorado.edu
Homepage: http://cechlab.colorado.edu/

functional group substitutions at specific sites; this "chemical genetics" can provide a detailed understanding of RNA active sites, or it can be used to append a wide variety of reporter groups to monitor RNA folding, conformational dynamics, and substrate binding.

Applying a Chemist's Viewpoint to Questions of RNA Mechanism

A common misperception about kinetic and thermodynamic measurements is that they are simply quantifying what is already apparent. In fact, the correct qualitative understanding of an RNA reaction often depends on making the right quantitative measurements. I learned this principle from Dan Herschlag when he was a postdoc in my group, so I choose two of many examples from Dan's work. The "ribozyme" described below is an RNA catalyst derived from the *Tetrahymena* self-splicing intron that can bind exogenous RNA substrates, cleave them by transesterification with a guanosine nucleophile, release the cleaved products, and repeat the catalytic cycle many times (Figure 1 a).

The first example started with a conundrum: RNA substrates that base-paired to the active site of the ribozyme with a single-base mismatch were cleaved substantially faster than those that bound with a perfect set of Watson–Crick base pairs. Why was the mismatched RNA a better substrate than the perfectly matched one? Did the mismatch somehow tweak the location of the scissile phosphate so that it was in a better orientation for nucleophilic attack? The answer turned out to be much simpler. Kinetic and thermodynamic analysis of the reaction showed that the rate-limiting step for cleavage of the matched RNA was release of the cleaved product from the active site.[15] Mismatches in the RNA actually decreased the rate of RNA binding and cleavage, but these effects were invisible when product release was rate-limiting; mismatches in the RNA–ribozyme pairing increased k_{off} and thereby accelerated the multiple-turnover reaction.

In the second example, graduate student Ben Young teamed with Dan Herschlag to investigate mutations in the ribozyme that shortened the RNA linker between the RNA substrate-binding site and the catalytic core.[16] Small deletions in the linker (for example, the −2 A ribozyme had two of three A nucleotides deleted) had dramatic effects, improving the turnover number and the sequence specificity of RNA cleavage. Rate constants for individual reaction steps were measured, revealing that the higher turnover and specificity both resulted from decreased affinity for the RNA substrate. The specificity analysis is shown in Figure 1 b. Thus, applying the sort of analysis typical of physical chemistry to the ribozyme reaction provided a logical, quantitative explanation of an otherwise puzzling phenomenon.

Chemical Synthesis Allows Chemical Genetics of RNA

A common set of questions arise for many RNAs that have an active role in biology. How

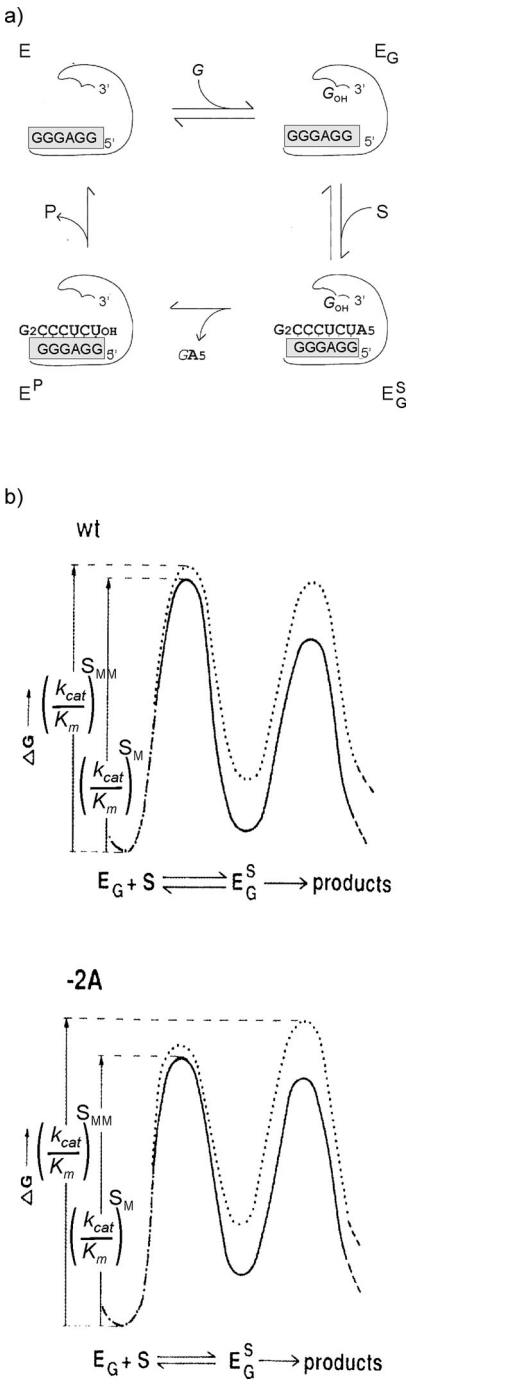

Figure 1. a) Cleavage of the "matched" RNA substrate S (GGCCCUCUAAAA) by the *Tetrahymena* ribozyme.[16] The ribozyme (thin line) binds guanosine triphosphate (G) to form the non-covalent E_G complex. It also binds S by Watson–Crick base-pairing to form the non-covalent ternary complex $E_G{}^S$. Cleavage of the RNA by transesterification gives two products: GA_5 is released quickly and GGCCCUCU remains bound in the intermediate E^P. Slow release of P restores the free ribozyme E. b) Mutant ribozyme has increased specificity for the matched substrate S_M ($GGCCCUCUA_5$) relative to the mismatched substrate S_{MM} ($GGCCC\underline{G}CUA_5$, where underlined base is mismatched with active site). Free energy reaction profiles under conditions of saturating G and subsaturating S_M (solid line) or S_{MM} (dotted line).[16] Wild-type (wt) ribozyme is rate-limited by substrate binding, so the activation barrier for cleavage of S_{MM} is only slightly higher than for S_M; there is only 10-fold sequence specificity. Mutant (−2A) ribozyme has a destabilized $E_G{}^S$ ground state, causing a change in rate-limiting step for cleavage of S_{MM} and giving 700-fold cleavage specificity for $S_M > S_{MM}$. (Reprinted from Ref. [16], Copyright (1991), with permission from Elsevier.)

does the RNA fold, and what is the folded structure? Are conformational dynamics important in the function? Where and how do small-molecule substrates bind to the RNA? If the RNA is catalytic, what is the mechanism? The goal of the chemist in such projects is to use the simplest chemistry suited for the question, which may be as straightforward as adding an activated ester to an amine on the RNA to form an amide bond. Because of the enormous size of many catalytic RNA molecules, semi-synthetic methods are very useful: a small chemically synthesized RNA fragment containing the designed modification is ligated together with longer fragments produced by in vitro transcription with a bacteriophage RNA polymerase.

I will give a few specific examples of ribozyme research done by chemically trained postdocs in my group over the years.

Anna Marie Pyle and graduate student Felicia Murphy used chemical genetics and thermodynamic analysis to identify the RNA substrate-binding site within the folded ribozyme.[17]

Joe Piccirilli in collaboration with the Caruthers lab built an oligonucleotide 7-mer ribozyme substrate containing a 3′-*S*-phosphorothiolate linkage at the cleavage site. A metal-ion specificity switch, with thiophilic Mn^{2+} or Zn^{2+} activating cleavage, indicated that the metal ion (unshaded in Figure 2a) contributes to catalysis by coordination to the 3′-oxygen atom in the transition state.[18] This provided direct evidence that

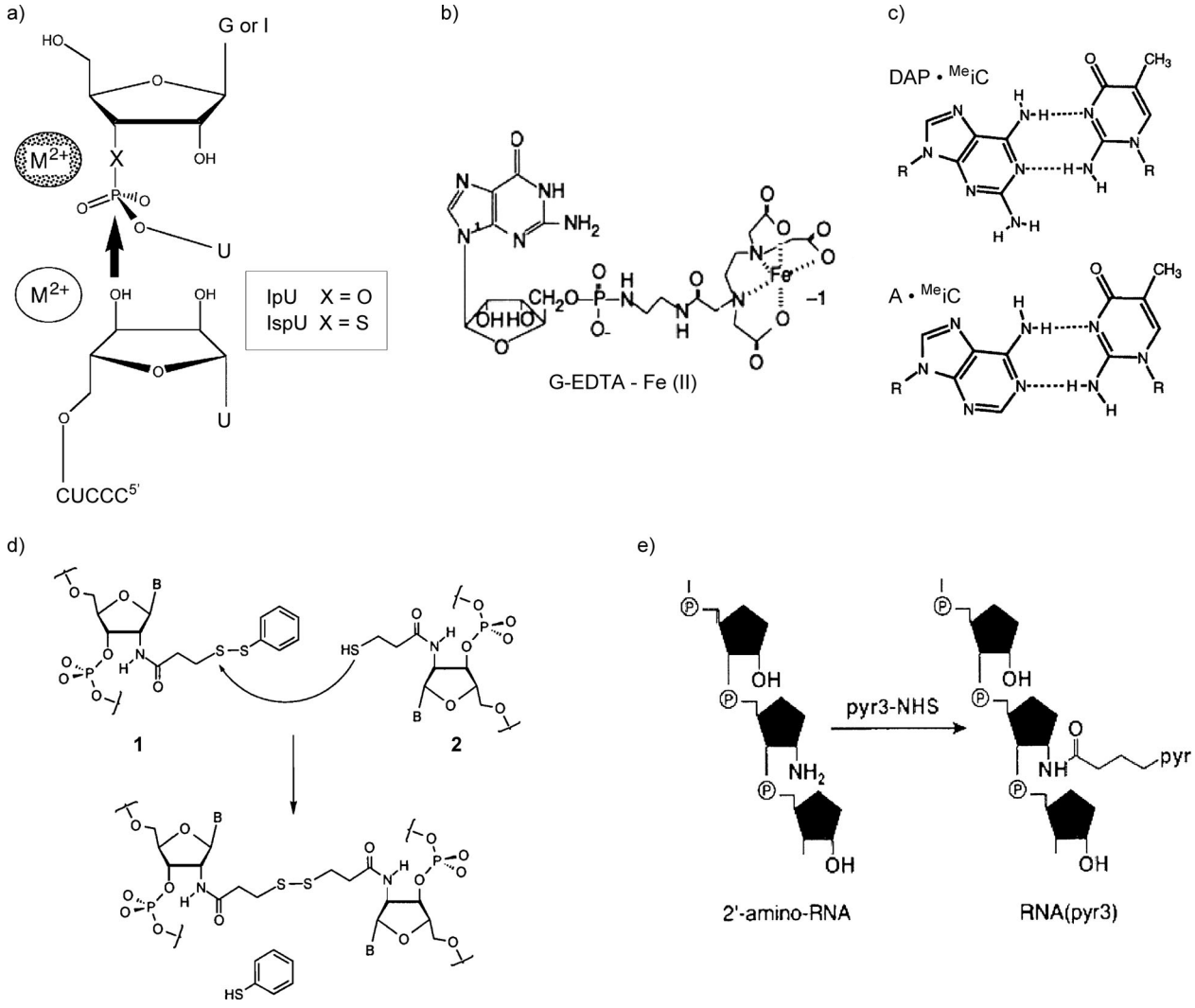

Figure 2. Synthetic chemistry to probe structure and mechanism of the *Tetrahymena* ribozyme. a) 3′-*S*-phosphorothiolate dinucleotide was reactive only in the presence of thiophilic metal ions (Cd^{2+} or Mn^{2+}), providing evidence for stabilization of the leaving group by the shaded metal ion (Mg^{2+} in standard reaction conditions).[19] b) Derivatization of the guanosine substrate with a metal chelator, binding of Fe^{II} and subsequent hydroxyl radical cleavage probed the tertiary structure around the G-binding site of the ribozyme.[20] c) Incorporation of base analogues into the ribozyme and its RNA substrate gave wobble base pairs that were isosteric with the natural G·U wobble pair and allowed measurement of the contribution of the amino group in the minor groove of the RNA helix.[21] d) Alkyl phenyl disulfide **1** and alkyl thiol **2** were incorporated at specific positions within the ribozyme; disulfide cross-linking by thiol–disulfide interchange allowed measurement of the dynamics of thermal motions within the large (310 nucleotide) ribozyme.[22] e) Oligoribonucleotide 15-mer with a 2′-amino substitution was derivatized with a pyrene (pyr) chromophore and then ligated to the remainder of a ribozyme domain to monitor RNA tertiary folding by fluorescence.[23]

the ribozyme is a metalloenzyme, mechanistically similar to certain protein enzymes.

Graduate student Lara Weinstein collaborated with the Cosstick lab to perform a similar analysis on the exon-ligation step of RNA splicing (reaction shown in Figure 2a).[19] They synthesized the requisite dinucleotide containing a 3′-*S*-phosphorothiolate and found that the ligation reaction required a thiophilic metal ion (Cd^{2+} or Mn^{2+}). This chemical genetic approach provided evidence for stabilization of the leaving group by a metal ion (shaded in Figure 2a), a role accomplished by Mg^{2+} in standard reaction conditions.

Jin-Feng Wang derivatized the guanosine substrate with a metal chelator (Figure 2b), added Fe^{II}, and obtained free-radical cleavage at a site brought into proximity by the RNA tertiary structure;[20] the identification of this portion of the ribozyme active site was later substantiated by X-ray crystallography.

Scott Strobel replaced the G·U wobble base pair at the RNA cleavage site with synthetic wobble base pairs that systematically altered functional groups (e.g., Figure 2c), and he concluded that the minor groove of the helix formed between the RNA substrate and its binding site is recognized by the ribozyme.[21]

Scott Cohen incorporated thiols and disulfides at specific sites in the ribozyme and monitored disulfide cross-linking (Figure 2d), revealing unexpected conformational dynamics in the ribozyme.[22]

Scott Silverman developed methods for site-specific labeling of large RNAs with a pyrene chromophore (Figure 2e) to follow RNA tertiary folding by fluorescence.[23]

Because this is in part a personal retrospective, I've emphasized examples from my own research group. However, many other RNA research groups have made similar contributions. Although this is not the place for a complete review, I note the seminal work of Harry Noller in using chemical approaches to map RNA structure and function within the active ribosome.[24]

In conclusion, being trained to think like a chemist is excellent preparation for RNA research. Furthermore, the ability to synthesize RNA with specific substitutions of unnatural nucleotides allows an RNA chemist to gain insights that would be difficult to achieve if restricted to studying natural RNA. In the other direction, RNA research provides numerous incentives to expand one's expertise in chemistry.

I dedicate this commentary to my PhD advisor, John Hearst, who taught me how physical chemistry could be applied to nucleic acids. I am grateful to the chemically trained students and postdoctoral fellows who applied their chemistry to catalytic RNA in my laboratory; they are listed here in chronological order, along with their current faculty position or company: Tan Inoue (Kyoto U.), Jim McSwiggen (RNA Society), John Latham (Alder Biopharmaceuticals), Sarah Woodson (Johns Hopkins), Jamie Williamson (Scripps), Dan Herschlag (Stanford), Anna Marie Pyle (Yale), Joe Piccirilli (U. Chicago), Jin-Feng Wang (Tianjin U.), Jennifer Doudna (UC Berkeley), Scott Strobel (Yale), Kevin Weeks (UNC Chapel Hill), Phil Bevilacqua (Penn State), Lara Weinstein Szewczak (Cell Press), Alex Szewczak (Merck), Scott Cohen (Children's Medical Research Inst., Sydney), Biliang Zhang (Guangzhou), and Scott Silverman (U. Illinois). I am also grateful to the great chemists who trained these postdocs: Leslie Orgel, Pete von Hippel, Chris Walsh, Don Crothers, Steve Boxer, Bill Jencks, Jackie Barton, Steve Benner, Nick Turro, Jack Szostak, Peter Dervan, Doug Turner, Alanna Schepartz, Peter Moore, Andy Myers, Ron Breslow, and Dennis Dougherty. Certainly I am just as grateful to the equal number of talented geneticists, molecular biologists, medical researchers and X-ray crystallographers who have inhabited the Cech lab, and perhaps the 125th anniversary of another journal will give me an opportunity to recognize their work.

Received: July 25, 2012
Published online: December 3, 2012

[1] a) T. R. Cech, A. J. Zaug, P. J. Grabowski, *Cell* **1981**, *27*, 487; b) K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147.
[2] B. L. Bass, T. R. Cech, *Nature* **1984**, *308*, 820.
[3] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849.
[4] H. Busch, R. Reddy, L. Rothblum, Y. C. Choi, *Annu. Rev. Biochem.* **1982**, *51*, 617.
[5] a) M. R. Lerner, J. A. Boyle, S. M. Mount, S. L. Wolin, J. A. Steitz, *Nature* **1980**, *283*, 220; b) R. Parker, P. G. Siliciano, C. Guthrie, *Cell* **1987**, *24*, 229; c) S. Valadkhan, A. Mohammadi, Y. Jaladat, S. Geisler, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11901.
[6] a) H. F. Noller, J. B. Chaires, *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 3115; b) H. F. Noller, *J. Bacteriol.* **1993**, *175*, 5297.
[7] P. B. Moore, T. A. Steitz, *Cold Spring Harb. Perspect. Biol.* **2010**, DOI: 10.1101/chsperspect.a003780.
[8] a) C. Tuerk, L. Gold, *Science* **1990**, *249*, 505; b) A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818; c) D. L. Robertson, G. F. Joyce, *Nature* **1990**, *344*, 467.
[9] R. R. Breaker, *Cold Spring Harb. Perspect.Biol.* **2010**, DOI: 10.1101/cshperspect.a003566.
[10] C. Yanofsky, *Nature* **1981**, *289*, 751.
[11] a) R. Lee, R. Feinbaum, V. Ambros, *Cell* **1993**, *75*, 843; b) B. Wightman, I. Ha, G. Ruvkun, *Cell* **1993**, *75*, 855; c) A. J. Hamilton, D. C. Baulcombe, *Science* **1999**, *286*, 950.
[12] J. T. Lee, *Cold Spring Harb.Perspect. Biol.* **2010**, DOI: 10.1102/chsperspect.a003749.
[13] M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, E. S. Lander, *Nature* **2009**, *458*, 223.
[14] T. R. Cech, *Angew. Chem.* **2000**, *112*, 34; *Angew. Chem. Int. Ed.* **2000**, *39*, 34.
[15] D. Herschlag, T. R. Cech, *Biochemistry* **1990**, *29*, 10172.
[16] B. Young, D. Herschlag, T. R. Cech, *Cell* **1991**, *67*, 1007.
[17] A. M. Pyle, F. L. Murphy, T. R. Cech, *Nature* **1992**, *358*, 123.
[18] J. A. Piccirilli, J. S. Vyle, M. H. Caruthers, T. R. Cech, *Nature* **1993**, *361*, 85.
[19] a) L. B. Weinstein, D. J. Earnshaw, R. Cosstick, T. R. Cech, *J. Am. Chem. Soc.* **1996**, *118*, 10341; b) L. B. Weinstein, B. C. N. M. Jones, R. Cosstick, T. R. Cech, *Nature* **1997**, *388*, 805.
[20] J.-F. Wang, T. R. Cech, *Science* **1992**, *256*, 526.
[21] S. A. Strobel, T. R. Cech, *Science* **1995**, *267*, 675.
[22] S. B. Cohen, T. R. Cech, *J. Am. Chem. Soc.* **1997**, *119*, 6259.
[23] S. K. Silverman, T. R. Cech, *Biochemistry* **1999**, *38*, 14224.
[24] a) D. Moazed, H. F. Noller, *Cell* **1986**, *47*, 985; b) S. Joseph, H. F. Noller, *EMBO J.* **1996**, *15*, 910.

