RNA Folding

A Packing-Density Metric for Exploring the Interior of Folded RNA Molecules**

*Jason P. Schwans, Nan-Sheng Li, and Joseph A. Piccirilli**

RNA molecules fold into intricate three-dimensional architectures that perform important biological functions. These complex architectures often contain coaxially stacked helices that adopt specific spatial arrangements through tertiary interactions between distinctly structured, complementary surfaces. These interactions result in the formation of expansive close-packed interfaces.[1] Within these interfaces, atoms acquire distinct packing environments that together uniquely specify an RNA fold. To define these packing environments biochemically, we exploited a series of nucleotide analogues with 2′-substituents that span a range of molecular volumes (Figure 1). We used nucleotide analogue interference mapping (NAIM)[2] to examine the effect of these volume changes at every residue within the structurally well-defined, independently folding P4-P6 RNA domain.[3] Our results show that these analogues together provide a sensitive metric for the packing density of the atoms surrounding the 2′-hydroxy groups in an RNA molecule.

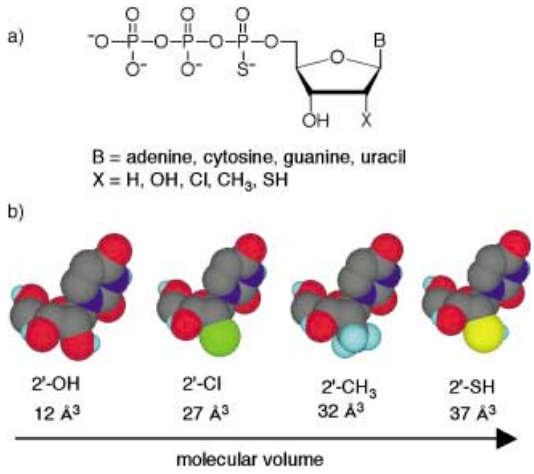

Figure 1. A series of nucleotide analogues that span a range of molecular volumes. A) Nucleotide-α-thiotriphosphates for NAIM analysis. B) Space-filling representations of the uridine analogues. The molecular volume of the 2′-substituent ranges from 12 to 37 Å^3. Molecular volumes were calculated by using the approach described by Connolly[4] and are consistent with volumes calculated from crystallographic data.[5]

We synthesized 2′-chloro, 2′-methyl, and 2′-sulfanylnucleoside analogues of adenosine, cytidine, guanosine, and uridine and converted these analogues into the corresponding α-thiotriphosphates. We have reported the synthesis of 2′-sulfanylnucleotide analogues previously.[6] The 2′-deoxy-2′-methylnucleoside analogues were synthesized by glycosylation of methyl 2-α-acetoxymethyl-3,5-di-*O*-(4-chlorobenzyl)-2-deoxy-α-D-ribofuranoside, followed by deoxygenation and deprotection.[7] The procedure described by Verheyden and Moffatt served as a guide for the synthesis of 2′-chloro-2′-deoxyuridine.[8] Displacement of the corresponding 2′-β-*O*-triflate esters with lithium chloride gave 2′-chloro-2′-deoxycytidine, adenosine, and guanosine.[9] We prepared the α-thiotriphosphate derivatives of the chloro- and methylnucleosides as described previously for the sulfanylnucleosides.[6]

[*] J. P. Schwans, N.-S. Li, Prof. J. A. Piccirilli
Howard Hughes Medical Institute
The University of Chicago
Department of Biochemistry and Molecular Biology
and Department of Chemistry
5841 S. Maryland Avenue, MC1028, Chicago, Il 60637 (USA)
Fax: (+1) 773-702-0271
E-mail: jpicciri@midway.uchicago.edu

[**] We thank Rui Sousa for the gift of the Y639F T7 RNA polymerase plasmid, Kara Juneau and Tom Cech for the ΔC209 P4-P6 plasmid, Cecilia Cortez and Shirshendu Deb for assistance with nucleoside synthesis, Joe Olvera for technical assistance, and Cheryl Small for assistance with manuscript preparation. We also thank members of the Piccirilli group and the reviewers for their comments on the manuscript. N.S.L. is a Research Specialist, and J.A.P. is an Associate Investigator of the Howard Hughes Medical Institute.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

 DOI: 10.1002/ange.200353575

To establish a metric for RNA packing density, we used interference mapping[2] to determine the influence of these analogues on the folding of ΔC209 P4-P6, an independently folding RNA domain derived from the *Tetrahymena* group I intron.[3] This domain serves as an excellent model system with which to explore the physical basis of analogue interference. Crystallographic analyses have revealed the domain structure,[3b,c] which consists of two sets of coaxially stacked helices that pack against each other to form a compact tertiary architecture. Biochemical analyses have allowed the thermodynamics and kinetics of folding of this domain to be defined.[10] Divalent magnesium ions induce a two-state tertiary folding transition that renders the folded molecules more compact than the unfolded molecules.[3a] Nondenaturing gel electrophoresis resolves these states and allows isolation (selection) of folded species from a population.[6,11]

We incorporated each nucleotide α-thiotriphosphate into ΔC209 P4-P6 at random by in vitro transcription with Y639F T7 RNA polymerase.[12] As described previously for the 2′-sulfanylnucleotides,[6] the polymerase efficiently and evenly incorporates the 2′-chloro- and 2′-methylnucleotides into the RNA to generate populations with approximately 1–5 analogue substitutions per transcript. Folded species were selected from the population by nondenaturing gel electrophoresis. Treatment with I_2 induced cleavage of the RNA backbone at sites containing phosphorothioate linkages [13] and thereby revealed the locations of the nucleotide analogues. Gaps in the I_2 cleavage ladder of the folded molecules compared to those of the unselected population indicate residues at which an analogue induces interference.[14] To achieve maximum sensitivity to folding interference without compromising the signal, we conducted our experiments at 0.45 mM Mg^{2+}, the concentration corresponding to the midpoint of the ΔC209 P4-P6 folding transition.[3c,6] Under these conditions, the selection assay detects interferences arising from as little as 0.2 kcal mol^{-1} destabilization.[6]

We observed little interference in simple duplex regions of ΔC209 P4-P6, which indicates that the minor groove of A-form helical RNA readily accommodates our atomic modifications. The interferences occur primarily in single-stranded regions and span the entire domain. When mapped onto the tertiary structure, the interferences cluster tightly in the three most densely packed regions of the molecule: the hinge region, the A-rich bulge/trihelical junction, and the tetraloop/receptor.[6] The residues in these regions frequently engage in hydrogen-bonding interactions through their 2′-hydroxy groups, experience significant surface area burial, and/or exhibit strong deviations from A-form helical geometry.[3b,c]

To explore the physical basis of these interferences, we obtained a 2′-deoxynucleotide (2′-H) interference map (Figure 2a) for comparison. Deoxynucleotide interference functionally distinguishes residues bearing dispensable (*nonfunctional*) 2′-hydroxy groups from those bearing *functional* 2′-hydroxy groups that provide a significant energetic contribution to folding. All deoxynucleotide interferences in ΔC209 P4-P6 occur at residues whose 2′-hydroxy groups engage in hydrogen bonds, as inferred from the crystal structure.[3b,c] The remaining residues in ΔC209 P4-P6 bear nonfunctional or dispensable 2′-hydroxy groups whose removal has little or no energetic effect on folding. Residues bearing functional hydroxy groups exhibit 2′-SH/2′-CH_3/2′-Cl interference, con-

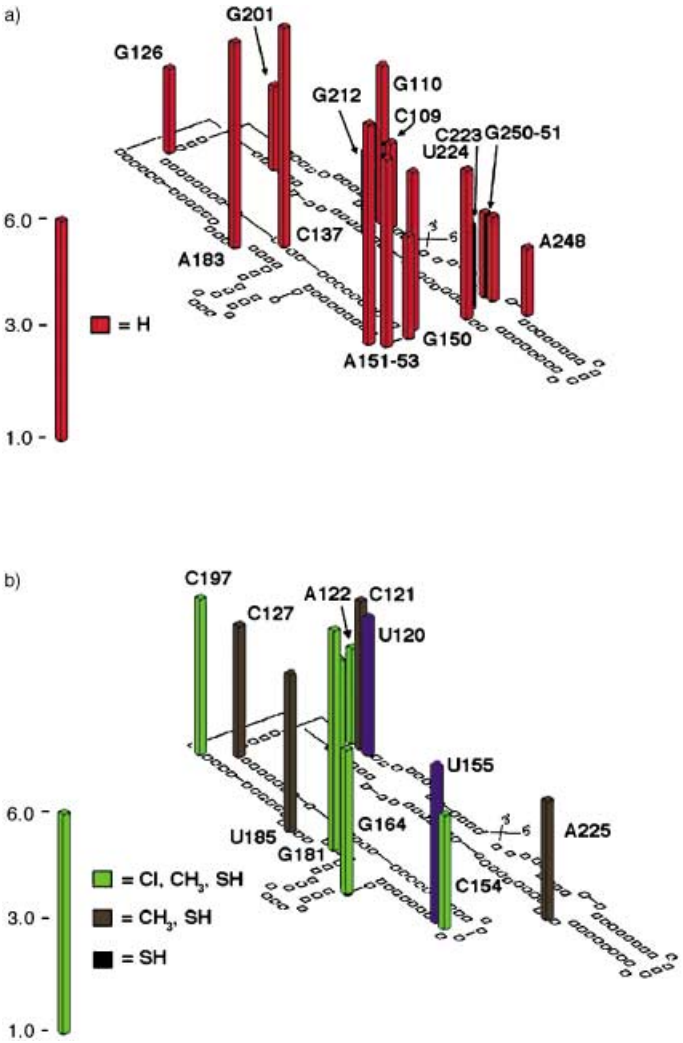

Figure 2. Interference histograms mapped onto the ΔC209 P4-P6 secondary structure. A) Deoxynucleotide interference distinguishes residues whose hydroxy groups engage in tertiary interactions from those that bear nonfunctional (dispensable) hydroxy groups. B) Superimposed interference maps for 2′-Cl, 2′-CH_3, and 2′-SH nucleotides. For clarity, the map shows only interferences at residues bearing nonfunctional hydroxy groups (no 2′-H interference). We conducted the folding selection at 0.45 mM Mg^{2+}, which corresponds to the midpoint of the ΔC209 P4-P6 folding transition,[3c,6] and calculated interference values as described by Ryder et al.[2e] For residues at which more than one analogue interferes, the histogram shows the maximum interference value observed. Residues that bear functional hydroxy groups generally exhibit the 2′-Cl/2′-CH_3/2′-SH interference profile. Sites at which one of the analogues interferes always also exhibit interference when an analogue of larger volume is present, such that three interference profiles emerge (green, brown, and blue bars). The data are summarized in Table 1.

sistent with our expectation that the analogues have little or no capacity to mediate tertiary interactions.[15] The residues bearing nonfunctional hydroxy groups therefore provide the basis set of microenvironments with which to define a metric for packing density. Among these residues, the number of interferences increases as the analogue volume increases; the interferences induced by a given analogue comprise a subset of the interferences induced by an analogue of larger volume (Figure 2b, Table 1). These results suggest that residues bearing nonfunctional hydroxy groups in ΔC209 P4-P6 have a volume threshold above which RNA folding is destabilized.

Table 1: ΔC209 P4-P6 interference profiles at 0.45 mM $MgCl_2$.

Residue	2′-Cl Int	2′-CH_3 Int	2′-SH Int
A122	X	X	X
C154	X	X	X
G164	X	X	X
G181	X	X	X
C197	X	X	X
G200	X	X	X
C121	–	X	X
C127	–	X	X
U185	–	X	X
A225	–	X	X
U120	–	–	X
U155	–	–	X

We used the crystal structure of ΔC209 P4-P6[3c] to determine the distances between the nonfunctional hydroxy groups and their nearest neighbors (Figure 3). Qualitatively, the interference profiles correlate with the number and closeness of neighboring atoms surrounding the hydroxy group. Residues for which all three analogues exhibit interference have the most densely packed hydroxy groups (Figure 3, left column); multiple close contacts usually occur with at least one atom residing within 2.65 Å of the hydroxy group. Residues that exhibit the 2′-SH/2′-CH_3 interference profile have less tightly packed hydroxy groups and usually only a single close contact within 3.1 Å (Figure 3, middle column). Residues that exhibit only 2′-SH interference have sparsely packed hydroxy groups whose closest atomic neighbors are between 3.2 and 3.5 Å away (Figure 3, right column). Residues that exhibit no interference have no close contacts. These observations suggest that the interference mapping volume thresholds reflect the spatial environment of the hydroxy groups in the folded state of ΔC209 P4-P6.

To define the space surrounding the hydroxy groups quantitatively, we introduced molecular volume changes at the hydroxy group sites and calculated the resulting change in van der Waals energy by using molecular mechanics.[17] Figure 4 shows volume–energy dependencies for five representative positions in ΔC209 P4-P6. As the molecular volume increases, the calculated energy rises, which reflects the increasing van der Waals repulsion caused by steric overlap. Sites that exhibit 2′-Cl/2′-CH_3/2′-SH interference display the greatest sensitivity to molecular volume. Sites that exhibit only 2′-SH interference have repulsive energy only after the molecular volume exceeds 35 $Å^3$. Sites that exhibit 2′-CH_3/2′-SH interference display an intermediate sensitivity and exert significant repulsion only when the molecular volume exceeds

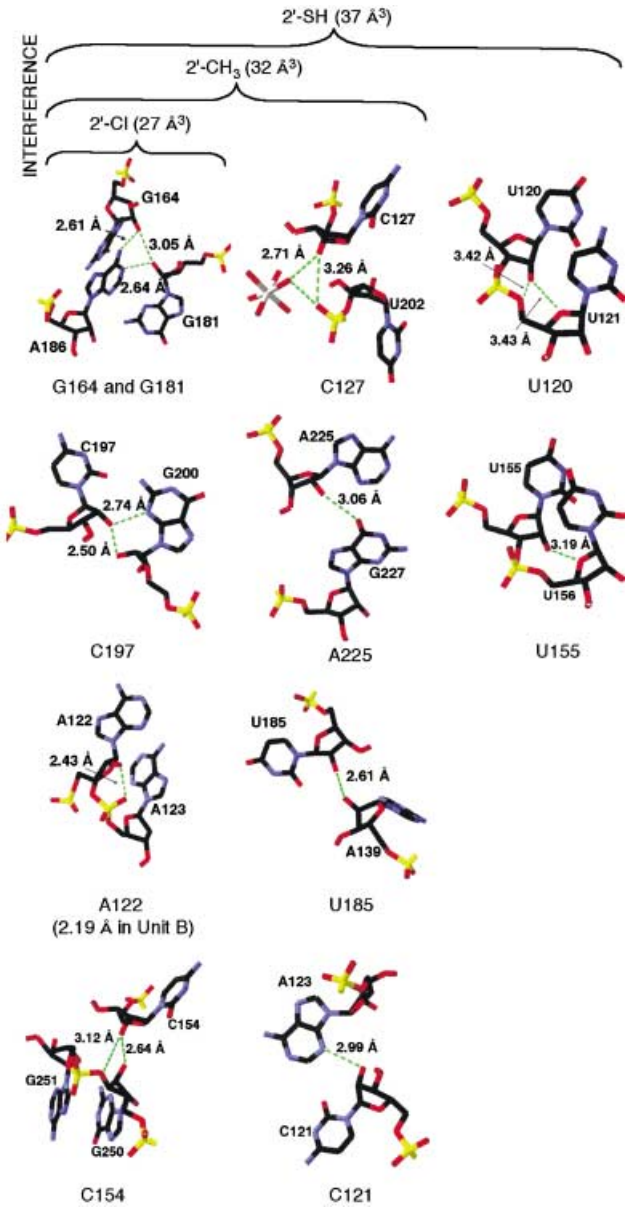

Figure 3. Contacts with neighboring atoms and interference profiles for nonfunctional 2′-hydroxy groups in ΔC209 P4-P6. Residues that exhibit 2′-SH interference have hydroxy groups whose closest atomic neighbors reside within 3.5 Å. Residues that exhibit interference in the presence of the intermediate-sized methyl group have hydroxy groups with closer neighbors (< 3.1 Å). Residues that exhibit interference caused by the smallest analogue (chlorine) have the most densely packed hydroxy groups with one or more close neighbors less than 2.65 Å away. We used Swiss–PDB Viewer software[16] to determine interatomic distances within the ΔC209 P4-P6 crystal structure.[3c] The atomic distances for unit A in the asymmetric unit are shown. Values for unit B are essentially the same as those for unit A, except that in unit B the 2′-oxygen atom of A122 is 2.19 Å from the phosphate oxygen atom of A123.

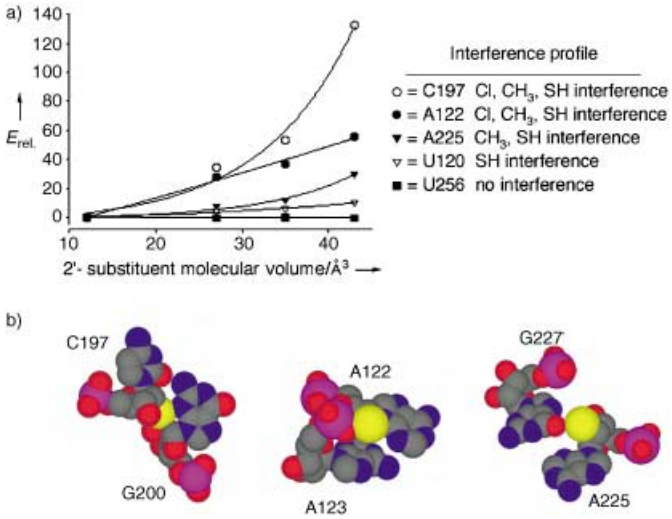

Figure 4. van der Waals repulsion determines the 2′-Cl/2′-CH_3/2′-SH interference profile. A) The dependence of van der Waals repulsion energy on molecular volume. The interference profile for a given residue matches its relative volume sensitivity. We used the Sybyl 6.9 molecular mechanics package[17] to calculate the change in van der Waals energy upon introduction of variously sized atoms at the 2′-position (O, Cl, S, Br). The calculation included all force potentials but ΔC209 P4-P6 was held rigid so that ΔE represents solely changes in van der Waals overlap. The smooth curves serve to guide the eye. B) Representative close-packed hydroxy groups and their local environment in space-filling representation. The models contain sulfur atoms (yellow) in place of the 2′-hydroxy groups.

30 Å^3. These trends indicate that the interference profiles reflect the volume–energy dependencies.

The multifunctional hydroxy group exerts its influence on RNA structure and function in multiple ways, for example, by mediating tertiary interactions through hydrogen bonding or metal-ion coordination, or by serving as a scaffold for the integral hydration network associated with RNA.[1] The 2′-hydroxy group may also make indirect contributions to structure and function as a consequence of its polar character and capacity to fill space, or its ability to withdraw electrons inductively or engender a ribonucleotide with a preference for the 3′-*endo* sugar conformation.[18] These effects apparently have no energetic significance for folding at residues bearing dispensable 2′-hydroxy groups (that is, residues that do not interfere upon 2′-deoxynucleotide substitution). Nevertheless, the nucleotide analogues described herein can "induce" ΔC209 P4-P6 folding interference when introduced at these locations. The analogues may act at several levels, which include, but are not necessarily limited to, 1) sugar conformational preferences, 2) hydrophobicity, and 3) molecular volume.

ΔC209 P4-P6 can be used as a model system to define the influence of atomic modifications on folding.[6] In addition to the analogues described herein, we have defined the influence of 2′-fluoro-, 2′-amino-, and 2′-methoxynucleotides on folding in the context of this RNA domain.[6,19] Only the 2′-fluoronucleotide analogues induce interference in accordance with their strong preference for the 3′-*endo* sugar conformation by destabilizing the tertiary structure when located at residues that populate the 2′-*endo* conformation in the folded RNA.[2d,19,20] We observe no link between interference and nucleoside analogues that preferentially populate the 2′-*endo* conformation.[6,19] These sugar pucker preferences evidently have little energetic influence on the ΔC209 P4-P6 tertiary folding transition, which was monitored by using the electrophoretic mobility shift assay. We did not observe a link between interference and the hydrophobic character of the introduced substituent, though we cannot rule out the possibility that hydrophobicity has some influence on tertiary folding stability.[18] Overlapping molecular volumes exert the most significant energetic force in biology, since the repulsive energy between two atoms increases exponentially with the inverse of the distance between the centers of the two atoms.[21] Our observation that ΔC209 P4-P6 interferences correlate with molecular volume is in agreement with this relationship.

In summary, we have synthesized 2′-chloro-, 2′-methyl-, and 2′-sulfanylnucleoside-α-thiotriphosphates and have demonstrated that T7 RNA polymerase efficiently incorporates these compounds into RNA without sequence bias, which renders interference mapping analysis possible. These nucleoside analogues span a narrow range of molecular volume ($\Delta V = 10$ Å^3) and together provide a sensitive measure of the spatial environment (the packing density) of a 2′-hydroxy group within a structured RNA molecule. Many residues within RNA bear nonfunctional hydroxy groups that reside within the extensive packing interfaces formed by tertiary folding. Our packing density metric (PDM) defines these interfaces biochemically. PDM analysis may help to evaluate the functional validity of crystallographic and biochemical models of RNA structure and may reveal regions within RNA that undergo structural changes whilst carrying out their function.

Received: December 18, 2003
Revised: March 15, 2004 [Z53575]

Keywords: hydrogen bonds · interfaces · nucleotides · RNA structures · RNA

[1] See, for example: a) S. A. Strobel, J. A. Doudna, *Trends Biochem. Sci.* **1997**, *22*, 262–266; b) A. R. Ferre-D'Amare, J. A. Doudna, *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 57–73; c) T. Hermann, D. J. Patel, *J. Mol. Biol.* **1999**, *294*, 829–849; d) R. T. Batey, R. P. Rambo, J. A. Doudna, *Angew. Chem.* **1999**, *111*, 2472–2491; *Angew. Chem. Int. Ed.* **1999**, *38*, 2326–2343; e) E. A. Doherty, J. A. Batey, B. Masquida, J. A. Doudna, *Nat. Struct. Biol.* **2001**, *8*, 339–343; f) D. J. Battle, J. A. Doudna, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11676–11681.

[2] a) R. B. Waring, *Nucleic Acids Res.* **1989**, *17*, 10281–10293; b) E. L. Christian, M. Yarus, *J. Mol. Biol.* **1992**, *228*, 743–758; c) S. A. Strobel, K. Shetty, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2903–2908; d) L. Ortoleva-Donnelly, A. A. Szewczak, R. R. Gutell, S. A. Strobel, *RNA* **1998**, *4*, 498–519; e) S. P. Ryder, L. Ortoleva-Donnelly, A. B. Kosek, S. A. Strobel, *Methods Enzymol.* **2000**, *317*, 92–109, and references cited therein.

[3] a) F. L. Murphy, T. R. Cech, *Biochemistry* **1993**, *32*, 5291–5300; b) J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden,

C. E. Kundrot, T. R. Cech, J. A. Doudna, *Science* **1996**, *273*, 1678 – 1685; c) K. Juneau, E. Podell, D. J. Harrington, T. Cech, *Structure* **2001**, *9*, 221 – 231.
[4] M. L. Connolly, *J. Am. Chem. Soc.* **1985**, *107*, 1118 – 1124.
[5] H. Ammon, *Struct. Chem.* **2001**, *12*, 205 – 212.
[6] J. P. Schwans, C. N. Cortez, J. O. Olvera, J. A. Piccirilli, *J. Am. Chem. Soc.* **2003**, *125*, 10012 – 10018.
[7] The synthesis of the 2′-deoxy-2′-methylnucleosides will be reported elsewhere.
[8] J. P. H. Verheyden, J. G. Moffatt, *J. Org. Chem.* **1972**, *37*, 2289 – 2299.
[9] See the Supporting Information for the synthesis of 2′-chloro-2′-deoxycytidine, adenosine, and guanosine, and characterization of an RNA transcript containing a single 2′-chloro-2′-deoxyuridine residue.
[10] a) S. K. Silverman, T. R. Cech, *Biochemistry* **1999**, *38*, 8691 – 8702; b) S. K. Silverman, M. L. Deras, S. A. Woodson, S. A. Scaringe, T. R. Cech, *Biochemistry* **2000**, *39*, 12465 – 12475; c) S. K. Silverman, T. R. Cech, *RNA* **2001**, *7*, 161 – 166.
[11] a) J. H. Cate, R. L. Hanna, J. A. Doudna, *Nat. Struct. Biol.* **1997**, *4*, 553 – 558; b) S. Basu, S. A. Strobel, *RNA* **1999**, *5*, 1399 – 1407.
[12] R. Sousa, R. Padilla, *EMBO J.* **1995**, *14*, 4609 – 4621.
[13] G. Gish, F. Eckstein, *Science* **1988**, *240*, 1520 – 1522.
[14] See the Supporting Information for representative ΔC209 P4-P6 sequencing ladders of transcripts containing phosphorothioates of guanosine, 2′-deoxyguanosine, and 2′-deoxy-2′-methylguanosine.
[15] Residue G150 bears a functional 2′-hydroxy group but exhibits no 2′-SH interference.[6]
[16] N. Guex, M. C. Peitsch, *Electrophoresis* **1997**, *18*, 2714 – 2723.
[17] M. Clark, R. D. Cramer III, N. V. Opdenbosch, *J. Comput. Chem.* **1989**, *10*, 982 – 1012.
[18] P. Gordon, R. Fong, S. K. Deb, N-S. Li, J. P. Schwans, J. A. Piccirilli, *Chem. Biol.* **2004**, *11*, 237 – 246, and references cited therein.
[19] J. P. Schwans, PhD Thesis, University of Chicago, Department of Chemistry, Chicago, **2003**.
[20] a) S. P. Ryder, S. A. Strobel, *J. Mol. Biol.* **1999**, *291*, 295; b) S. P. Ryder, S. A. Strobel, *Nucleic Acids Res.* **2002**, *30*, 1287.
[21] T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd ed. (Ed.: W. H. Freeman), New York, **1993**, pp. 140 – 148.

