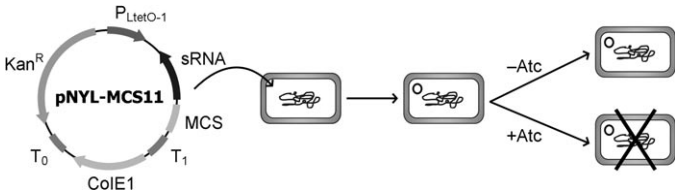
DOI: 10.1002/cbic.200800591

# Identification of a Toxic Peptide through Bidirectional Expression of Small RNAs

Wendy W. K. Mok,[a] Naveen K. Navani,[a, b] Courtney Barker,[a] Bobbijo L. Sawchyn,[a] Jimmy Gu,[a] Ranjana Pathania,[a, b] Rebecca D. Zhu,[a] Eric D. Brown,[a] and Yingfu Li*[a]

Research in the field of noncoding functional RNA sequences has flourished over the past two decades and has showcased the utility of these nucleic acids, which extends beyond their traditional roles as being the workhorses behind protein translation. A search through the intergenic regions of prokaryotic genomes has unveiled a class of important regulatory elements, known as small RNAs (sRNAs). As their names imply, these RNA sequences are relatively short, typically ranging from 50 to 400 nucleotides (nt) in length. In the *E. coli* genome alone, over 80 sRNA sequences have been discovered by using a combination of bioinformatics and experimental techniques, though the function of many of these sequences remains unknown.[1] Of the sRNAs that have been characterized to date, the majority of them interact with mRNA targets, thereby modulating their translation initiation or their stability upon binding.[2] Another subset of sRNAs can regulate the activity of protein targets or function as a part of a protein–sRNA complex.[2] Together, these tiny elements govern a number of bacterial stress response pathways[3] and they have been linked to pathogenicity in virulent species.[3–7] Due to the substantial role of sRNAs in maintaining cellular homeostasis and viability, disruption of their expression is expected to have deleterious effects on the cell.

Previously, screening for lethal or growth defective phenotypes upon the overexpression of antisense RNA sequences was used to identify essential genes in *Staphylococcus aureus*.[8,9] Here, we implemented a similar approach in an effort to develop a method to isolate sRNA sequences with critical regulatory functions in *E. coli*. We adopted a tetracycline-inducible system to regulate the expression of the sense and antisense sequences of twelve sRNAs. These sequences were cloned into pNYL-MCS11, a plasmid derived from pZE21-MCS1, which is described in an earlier study.[10] The optimized ribosome binding site from the plasmid was removed prior to cloning, such that the sense and antisense sRNAs would be transcribed but not translated when their expression is induced in the presence of anhydrotetracycline (Atc), an analogue of tetracycline that appears to be a more potent yet less toxic inducer.[11] Analyses were subsequently performed in *E. coli* strain DH5αZ1, which has been engineered to endogenously express the TetR repressor (Scheme 1).[10] Of the sRNAs that were screened, three have been extensively studied and have been shown to regulate different cellular processes (Table 1). These candidates were chosen in order to examine the physiological effects associated with their up- and down-regulation. The functions of the remaining nine sequences have not yet been characterized. As such, we wish to probe into their biological roles through this screen. Overexpression of all twelve sense sRNA sequences did not affect cell growth




**Scheme 1.** Overview of the screening approach adapted for the identification of sRNA sequences with critical functions. Sense and antisense sequences of twelve sRNAs were cloned into pNYL-MCS11 downstream of a tetracycline-inducible promoter. This construct was then transformed into *E. coli* by electroporation. Lethal phenotypes linked to the overexpression of the sequences was screened by inducing their synthesis with Atc.

**Table 1.** List of sRNA sequences screened.

| | sRNA | Length [nt] | Adjacent genes | Function | Refs. |
|---|---|---|---|---|---|
| 1 | SraD/MicA | 75 | luxS/gshA | regulates expression of the OmpA outer membrane protein | [12, 15] |
| 2 | SraI/RyhB | 90 | yhhX/yhhY | regulates expression of iron-storage and iron-containing proteins | [12, 16] |
| 3 | SsrA/ tmRNA/ PsrD | 363 | smpB/intA | rescue of stalled ribosomes in complex with SmpB | [12, 17] |
| 4 | RygC/T27/ QUAD1c | 140 | ygfA/serA | unknown | [12, 14] |
| 5 | SraB/ Pke20 | 169 | yceF/yceD | unknown | [12] |
| 6 | SraC/ RyeA/ Tpke79/ ISO091 | 249 | pphA/yebY | unknown | [12, 14] |
| 7 | Tpke11 | 89 | dnaK/dnaJ | unknown | [12, 18] |
| 8 | C0067 | 125 | yafT/yafU | unknown | [12, 19] |
| 9 | C0293 | 73 | lcd/ymfD | unknown | [12, 19] |
| 10 | C0299 | 79 | hlyE/umuD | unknown | [12, 19] |
| 11 | C0465 | 78 | tar/chew | unknown | [12, 19] |
| 12 | C0719 | 222 | glcA/glcB | unknown | [12, 19] |

[a] *W. W. K. Mok,[+] Dr. N. K. Navani,[+] C. Barker,[+] B. L. Sawchyn, J. Gu, Dr. R. Pathania, R. D. Zhu, Prof. Dr. E. D. Brown, Prof. Dr. Y. Li*
*Department of Biochemistry and Biomedical Sciences, McMaster University*
*1200 Main Street, W. Hamilton, ON L8N 3Z5 (Canada)*
*Fax: (+1) 905-522-9033*
*E-mail: liying@mcmaster.ca*

[b] *Dr. N. K. Navani,[+] Dr. R. Pathania*
*Present address:*
*Department of Biotechnology, Indian Institute of Technology*
*Roorkee 247667, Uttarakhand (India)*

[+] *These authors contributed equally to this work.*

(Figure 1, top panel). On the other hand, inducing the expression of their antisense counterparts led to the discovery of one sequence that can suppress cell growth when it is overexpressed (Figure 1, bottom panel). This sequence corresponded to the antisense of the RygC sRNA.

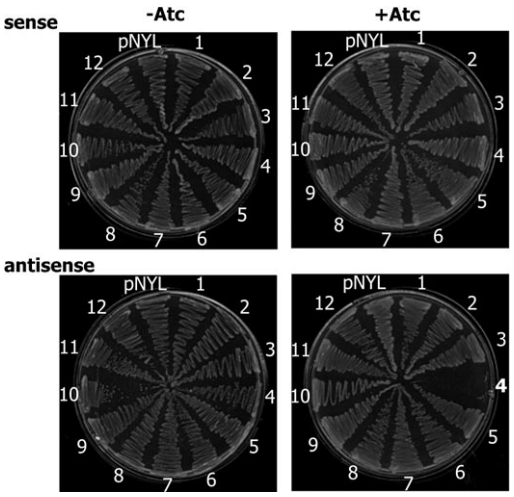



**Figure 1.** Effect of overexpressing sense and antisense variants of twelve sRNA sequences on the growth of *E. coli* strain DH5αZ1. Cells transformed with vectors encoding sense (top) and antisense (bottom) variants of 12 sRNAs were grown on solid media in the absence or presence of Atc. Through this screen, it was observed that the overexpression of the antisense variant of sequence 4, which corresponds to the RygC sRNA, led to growth inhibition.

Located in the intergenic region between *ygfA* and *serA* in the *E. coli* genome, *rygC* was shown to be conserved in *Salmonella typhimurium* and in *Shigella flexneri* through primary sequence alignments.[12] Secondary structure analyses indicated that it belongs to the QUAD 1 family of RNA structural motifs.[13] Northern hybridization analysis revealed that RygC is expressed in vivo and its expression is elevated during stationary phase growth and in cells growing in minimal media.[14] However, its function is still unknown. Herein, we seek to elucidate the function of RygC and the rationale behind antisense RygC (rRygC)-induced growth suppression.

In our initial screen, as shown in Figure 1, the consequences of antisense sRNA overexpression were qualitatively evaluated on solid media approximately 16 h following induction by Atc. Monitoring the growth of *E. coli* in liquid media demonstrated that growth inhibition caused by rRygC overexpression was not evident until 3 h after induction (Figure 2). The growth of cells carrying pNYL-MCS11 without the rRygC insert was unaffected in the presence of 200 ng mL$^{-1}$ of Atc. The delay in the onset of antisense-induced lethality might be a reflection of the time needed to express a sufficient quantity of rRygC, allowing it to disrupt the regulatory function of sense RygC or other physiological activities.

As a first step towards understanding the mechanism behind rRygC-induced growth suppression, we sought to determine the regions within its primary sequence that may be necessary for it to exude its inhibitory effect. This was achieved by truncating the DNA sequence encoding rRygC from the 5′- and 3′-ends, as well as introducing internal deletions (Figure 3 A). These truncated sequences were subsequently cloned into pNYL-MCS11. Inducing the expression of the sense counterparts of these truncated sequences was not shown to affect growth (Figure 3 B, top panel). Removal of the first 22 nt from the 5′-end of rRygC, which corresponded to the sequence complementary to RygC's putative transcription terminator, did not diminish its inhibitory effect, as indicated by truncated sequences 1 and 2 (Figure 3 B, bottom panel). Sequence 2 further demonstrates that the last 30 bases from the 3′-end in rRygC were dispensable. Overexpression of sequences harbouring deletions of the first 34 bases from the 5′-end, however, resulted

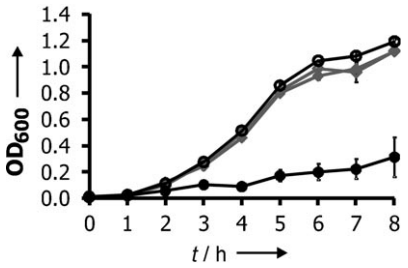



**Figure 2.** Growth of cells overexpressing antisense RygC. Cells carrying pNYL-MCS11 (denoted pNYL) and pNYL-MCS11 with the rRygC insert (rRygC) were grown overnight in Luria–Bertani (LB) medium with 50 μg mL$^{-1}$ of kanamycin before being subcultured into fresh LB-kanamycin media in the absence of Atc (−Atc) or in LB-kanamycin supplemented with 200 ng mL$^{-1}$ of Atc (+Atc). Growth of cells in each condition was monitored hourly for 8 h by measuring their absorbance at 600 nm. This experiment was carried out in triplicate. (◇ pNYL−Atc, ◆ pNYL+Atc, ○ rRygC−Atc, ● rRygC+Atc)

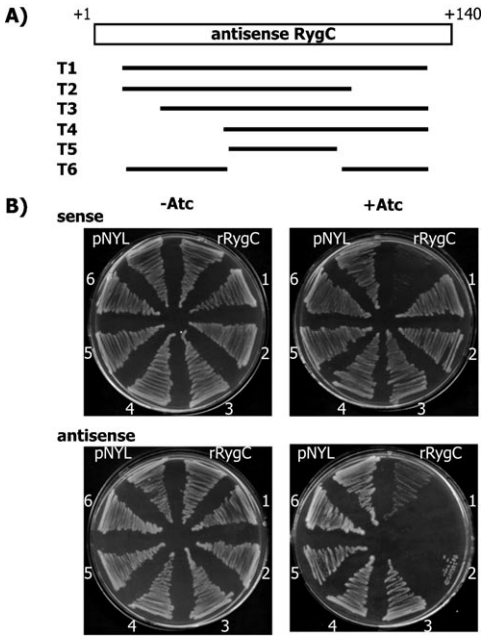



**Figure 3.** Truncation analysis of rRygC. A) Outline of truncated sequences 1–6. B) Effect of expressing the truncated sequences on the growth of *E. coli* DH5αZ1. It was observed that deletions from the 5′-end of antisense RygC abolished the deleterious effects associated with the overexpression of this sequence.

in the loss of lethal phenotype, as illustrated by truncated sequences 3 to 5. Since we were interested in probing for the nt that may be required for the rRygC transcript to elicit its growth-inhibitory function, we deleted 44 bases within its sequence along with the first 22 bases that were complementary to the proposed RygC terminator (sequence 6). Overexpression of this sequence did not interfere with the growth of *E. coli*, indicating that some of the deleted residues are vital for rRygC-associated lethality.

Further examination of the rRygC sequence revealed the presence of a ribosome binding site (RBS) 27 bases downstream of its transcription start site. This RBS is followed by two possible start codons. A stop codon is present 54 nt following the first start codon. Together, these elements suggest that the antisense RygC transcript encodes a peptide of 18 or 19 amino acids. In agreement with this model, the overexpression of truncated sequences 3, 4, and 5 was no longer detrimental to cell growth as deletions introduced into these sequences would have contributed to the removal of the RBS and/or start codon. As a result, the peptide can no longer be produced using these truncated transcripts. The internal deletions introduced into sequence 6 might have led to the removal of 33 nt, which encode eleven of the amino acids in the peptide in addition to the stop codon. Likewise, the deleterious effect to the cell when this sequence was induced was lost. To validate that rRygC encodes a peptide that compromises survival when it is abundant in the cell, we introduced mutations in the putative ribosome binding site and start codons by using site-directed mutagenesis (Figure 4A). rRygC affiliated lethality was retained when the second ATG site was mutated (M2). This indicates that this codon does not correspond to the start codon. However, when the first ATG codon (M1) and ribosome binding sites (M4) were mutated, the deleterious effect was abolished.

While we were preparing this manuscript, a study was published by Fozo and colleagues revealing that rRygC and the reverse complements of other members of the QUAD 1 family of sRNAs harbour an ORF (open reading frame) encoding a small hydrophobic peptide of 18 to 19 amino acids.[20] Through this study in which the protein encoded by rRygC was extensively characterized, it was confirmed that this protein, referred to as IbsC, was a toxin, whereas its corresponding sRNA was a transient antitoxin. Overexpression of IbsC was found to result in depolarization of membrane potential, suggesting that this toxin may act by disrupting membrane integrity. The results of this study corroborates our observations that the lethality associated with rRygC overexpression is attributed to the increased production of a toxic peptide.

Although we now understand the mechanism behind rRygC-induced growth suppression, the reason behind its synthesis in vivo and the mechanism behind its regulation remain unresolved. Since the RBS and start codon of this ORF overlaps with the transcribed region of sense RygC, we propose that the sRNA acts to suppress the translation of this peptide by occluding these translation initiation sites, thereby promoting cell survival (Figure 5A). Alternatively, similar to another subset of sRNAs that act on mRNA targets, RygC may destabilize the toxin-encoding transcript upon binding (Figure 5B).

The initial objective of this study was to devise a screening approach, based on antisense RNA knock-down technology, which would enable the identification of sRNA sequences critical for maintaining cellular integrity. We introduced sequences encoding the sense and antisense variants of twelve sRNAs in an RNA expression system and screened for sequences that can elicit death or growth defects when their expression is induced by Atc. Through this screen, we found that overexpressing the antisense counterpart of the RygC sRNA was inhibitory to cell growth. Closer examination of this sequence revealed that the inhibitory effect was not due to antisense knock-down of RygC. Rather, it was serendipitously found that this antisense transcript contains an ORF, which encodes a toxic peptide of 19 amino acids. We believe that our method can be widely used for the identification of potential sRNA or other genes with intriguing functions, which may be overlooked using conventional techniques.

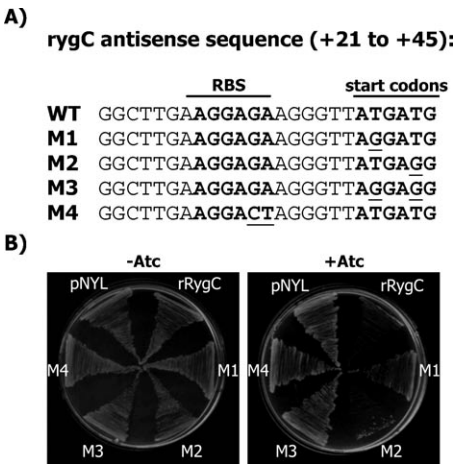



**Figure 4.** A) The predicted RBS and putative start codons encoded by rygC are illustrated above. Mutant sequences M1 to M4 contain point mutations in these sites. B) Inducing the expression of sequences M1, M3 and M4 using Atc was no longer toxic to *E. coli*. Sequence M2 retained its deleterious effect, suggesting that the second ATG codon does not correspond to the start codon.

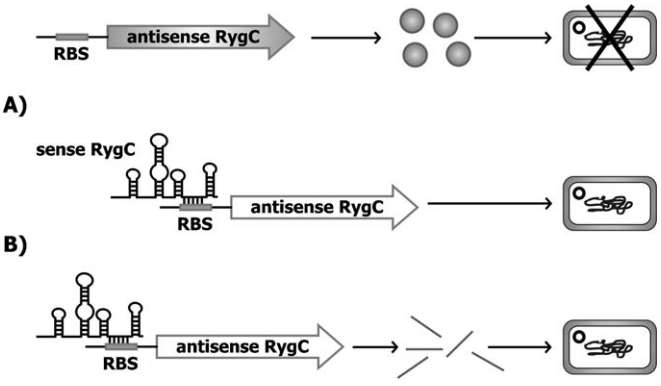



**Figure 5.** Hypothesis of rRygC regulation. A) rRygC encodes a toxic peptide, and the translation of this peptide is regulated by RygC. B) Interaction between the RygC sRNA and its antisense transcript leads to destabilization of the transcript and prevention of its translation.

## Experimental Section

**Oligonucleotides and Reagents**: All PCR primers and oligonucleotides used in this study were chemically synthesized by Integrated DNA Technologies (Coraville, IA, USA). The antibiotics used in this study, such as kanamycin and anhydrotetracycline (Atc), were purchased from Sigma–Aldrich. For molecular cloning, the High Fidelity PCR enzyme mix and T4 DNA ligase were purchased from Fermentas (Burlington, ON, Canada). Restriction enzymes were obtained through either Fermentas or New England Biolabs. For site-directed mutagenesis of the rRygC-encoding vector, PfuUltra High Fidelity DNA Polymerase and Dpn1 restriction enzyme from Stratagene were used to amplify the mutated plasmids and cleave the template vector, respectively. Plasmid Miniprep kits were purchased from either Qiagen or Promega.

**Plasmids and bacterial strains**: Molecular cloning steps, including PCR, restriction digestion, and ligation were conducted following established protocols provided by suppliers. Sense and antisense sRNA-encoding sequences used for the lethality screen were amplified from *E. coli* DH5αZ1 genomic DNA by PCR. Likewise, truncated sense and antisense *rygC* sequences 1–4 were amplified by PCR using full-length antisense *rygC* as template. Truncated sequences 5 and 6 were obtained by hybridizing complementary synthetic oligonucleotides. These PCR amplified and synthetically prepared sRNA sequences were cloned into EcoRI and BamHI sites in pNYL-MCS11, a derivative of pZE21-MCS1 (courtesy of H. Bujard). Prior to cloning, pZE21-MCS1 was digested with restriction enzyme EcoRI in order to remove its optimal ribosome binding site. A new multiple cloning site was subsequently introduced at the EcoRI site, which allowed for the replacement of the restriction sites that were lost upon RBS removal. Plasmids carrying sense and antisense sRNA sequences or truncated *rygC* variants were transformed into *E. coli* strain DH5αZ1, which endogenously expresses the tetracycline repressor, by electroporation.

**Lethality screens**: Transformants with plasmids containing sense and antisense sRNA-encoding inserts or truncated sense and antisense *rygC* variants were grown at 37 °C on solid Luria-Burtani (LB) medium supplemented with kanamycin (50 $\mu g\,mL^{-1}$) in the absence (for uninduced conditions) or presence (for induced conditions) of Atc (400 $ng\,mL^{-1}$). The effects of overexpressing the various sRNA or *rygC*-encoding constructs were evaluated following overnight growth.

***E. coli* growth curve**: Cells carrying pNYL-MCS11 or pNYL-MCS11 with an rRygC insert were grown overnight at 37 °C with shaking at 260 rpm in LB broth (3 mL) supplemented with kanamycin (50 $\mu g\,mL^{-1}$). Following overnight growth, the cells were diluted by 200-fold in fresh LB-kanamycin media (5 mL; for uninduced conditions) or in LB-kanamycin media (5 mL) supplemented with Atc (200 $ng\,mL^{-1}$; for induced conditions). These cultures were incubated at 37 °C with shaking for 8 h. Cells growth was monitored hourly by measuring the optical density at 600 nm ($OD_{600}$) for 0.2 mL of each culture using a SpectraMax Plus 384 plate reader (Molecular Devices, Sunnyvale, CA, USA). The assay was conducted in triplicate.

**Mutagenesis of antisense *rygC***: The ribosome binding site, start codon(s), and stop codon of rRygC were mutated by site-directed mutagenesis following supplier-provided protocols (Stratagene). Briefly, forward and reverse mutagenic primers of 34–41 nt in length were designed, such that they would hybridize to the region that is to be changed and introduce point mutations to these sites. The rRygC-encoding vector was then amplified using these primers along with the PfuUltra High Fidelity DNA Polymerase (Stratagene). Following amplification, the methylated template plasmid was cleaved following two rounds of digestion using Dpn1 restriction enzymes (Stratagene). The reaction mixtures were subsequently transformed into *E. coli* strain XL-1 blue competent cells by electroporation. Plasmids isolated from the transformants were sequenced (Mobix Lab, McMaster Univeristy) in order to identify those carrying the desired mutations. The mutated plasmids were then transformed into *E. coli* DH5αZ1 cells. The lethality attributed to the overexpression of these sequences was screened on solid LB-kanamycin media in the absence or presence of Atc (400 $ng\,mL^{-1}$) following the aforementioned protocol.

## *Acknowledgements*

*We thank Professor Hermann Bujard for the gift of plasmids and* E. coli *DH5αZ1 strain. We thank Pascale Libront Swanson for her help in constructing pNYL-MCS11 and the Li Lab members for helpful discussions. This project is partially funded by the Ontario Premier Research Excellence Award and Canada Research Chairs Program. E.D.B. and Y.L. are Canada Research Chairs. W.M. holds an NSERC graduate student scholarship.*

**Keywords:** antisense RNA · mutagenesis · non-coding RNA · RNA · small RNA · toxic peptide

[1] J. Vogel, E. G. Wagner, *Curr. Opin. Microbiol.* **2007**, *10*, 262.
[2] G. Storz, S. Altuvia, K. M. Wassarman, *Annu. Rev. Biochem.* **2005**, *74*, 199.
[3] S. Gottesman, *Trends Genet.* **2005**, *21*, 399.
[4] G. Padalon-Brauch, R. Hershberg, M. Elgrably-Weiss, K. Baruch, I. Rosenshine, H. Margalit, S. Altuvia, *Nucleic Acids Res.* **2008**, *36*, 1913.
[5] D. H. Lenz, K. C. Mok, B. N. Lilley, R. V. Kulkarni, N. S. Wingreen, B. L. Bassler, *Cell* **2004**, *118*, 69.
[6] V. Pfeiffer, A. Sittka, R. Tomer, K. Tedin, V. Brinkmann, J. Vogel, *Mol. Microbiol.* **2007**, *66*, 1174.
[7] A. Toledo-Arana, F. Repoila, P. Cossart, *Curr. Opin. Microbiol.* **2007**, *10*, 182.
[8] R. A. Forsyth, R. J. Haselbeck, K. L. Ohlsen, R. T. Yamamoto, H. Xu, J. D. Trawick, D. Wall, L. Wang, V. Brown-Driver, J. M. Froelich, K. G. C, P. King, M. McCarthy, C. Malone, B. Misiner, D. Robbins, Z. Tan, Z. Y. Zhu Zy, G. Carr, D. A. Mosca, C. Zamudio, J. G. Foulkes, J. W. Zyskind, *Mol. Microbiol.* **2002**, *43*, 1387.
[9] Y. Ji, B. Zhang, S. F. Van Horn, P. Warren, G. Woodnutt, M. K. Burnham, M. Rosenberg, *Science* **2001**, *293*, 2266.
[10] R. Lutz, H. Bujard, *Nucleic Acids Res.* **1997**, *25*, 1203.
[11] J. Degenkolb, M. Takahashi, G. A. Ellestad, W. Hillen, *Antimicrob. Agents Chemother.* **1991**, *35*, 1591.
[12] R. Hershberg, S. Altuvia, H. Margalit, *Nucleic Acids Res.* **2003**, *31*, 1813.
[13] K. E. Rudd, *Res. Microbiol.* **1999**, *150*, 653.
[14] K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, *Genes Dev.* **2001**, *15*, 1637.
[15] K. I. Udekwu, F. Darfeuille, J. Vogel, J. Reimegard, E. Holmqvist, E. G. Wagner, *Genes Dev.* **2005**, *19*, 2355.
[16] E. Masse, S. Gottesman, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 4620.
[17] P. W. Haebel, S. Gutmann, N. Ban, *Curr. Opin. Struct. Biol.* **2004**, *14*, 58.
[18] E. Rivas, R. J. Klein, T. A. Jones, S. R. Eddy, *Curr. Biol.* **2001**, *11*, 1369.
[19] B. Tjaden, R. M. Saxena, S. Stolyar, D. R. Haynor, E. Kolker, C. Rosenow, *Nucleic Acids Res.* **2002**, *30*, 3732.
[20] E. M. Fozo, M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, G. Storz, *Mol. Microbiol.* **2008**.

Received: September 3, 2008
Published online on December 17, 2008