Short Communication

Prediction of structural homologs to functional RNAs involved in determination of life span of *Paramecium tetraurelia*

Hiroyuki TANABE, [1,*,†] and Shu-Yun LE[2,*]

[1]*Laboratory of Biochemistry, Faculty of Agriculture, Kinki University, Nara 631-8505, Japan*
[2]*Center for Cancer Research Nanobiology Program, NCI Center for Cancer Research, National Cancer Institute, NIH, Frederick, MD 21702, USA*

SUMMARY

The MS2 gene of *Paramecium tetraurelia* is supposed to be intimately implicated in the programmed association between sexual reproduction and life span and in the eventual setting of its maximum life span. The transcript of this gene is presumably an mRNA-like non-coding RNA, the function of which is evolutionarily conserved in its specific secondary structure. Therefore, characterizing the secondary structure of MS2 RNA would yield important insights about the molecular mechanism for the above cellular process that is shared among all of the organisms possessing limited longevity. In this study, it was found that *P. tetraurelia* has a close homolog of MS2 which undergoes the same regulation of expression as with MS2. We computationally analyzed the RNA secondary structures of MS2 and the homolog and detected four types of well-ordered structures that are both significantly more ordered and thermodynamically stable than anticipated at random. Our database searches revealed that these consensus structural features are conserved from other protozoan species to mammals. These results may suggest the functional importance of the distinct RNA structures in relation to the cellular program for organismal life span.

Although considerable advancements are acknowledged in the studies to delay aging and thereby prolong life of model organisms (de Magalhaes, 2005; Guarente, 2005), the underlying mechanism for determination of the species-specific maximum life span (Takagi, 1999) is scarcely understood. The ciliated protozoan *Paramecium* has an intrinsic length of clonal life span which corresponds to organismal life span in multicellular organisms (Sonneborn, 1954). For the final purpose of identifying the *jumyo* gene, the causal determinant of the clonal life span of *Paramecium tetraurelia* (Takagi et al., 1989), we examined differences in gene expression between wild-type stock 51 and its mutant d4-SL4 which has an extremely short clonal life span (Takagi et al., 1987), and isolated three genes that showed expression profiles dependent on clonal age (Tanabe et al., 2002). Of the three genes, MS2, cloned as a

*The two authors contributed equally to this study.
[†]Corresponding author
Fax: +81-742-43-8976
E-mail: htanabe@nara.kindai.ac.jp
Received: 18 April 2006; Accepted: 12 June 2006.

gene expressed predominantly in the short-lived mutant, was gradually activated in the wild-type paramecia as their clonal age increased. *Paramecium tetraurelia* terminates its clonal life span either in clonal death or in causing autogamy, a form of sexual reproduction consisting of meiosis and self-fertilization (Sonneborn, 1974). In the course of another strategy for identification of the *jumyo* gene, molecular understanding of the mechanism of autogamy, we showed that the MS2 gene became up-regulated in the wild-type stock through the pathway of autogamy (Tanabe et al., 2003). The closely parallel transcriptional patterns of MS2 observed during clonal aging and during autogamy suggest that the two events may share a certain genetic process for timing the clonal life span, in which MS2 is functionally participated. Thus, MS2 is expected to provide the first molecular basis of the essential connection between reproduction and mortality (Takagi, 1999) and, in turn, the genetic program to delimit the organismal life span.

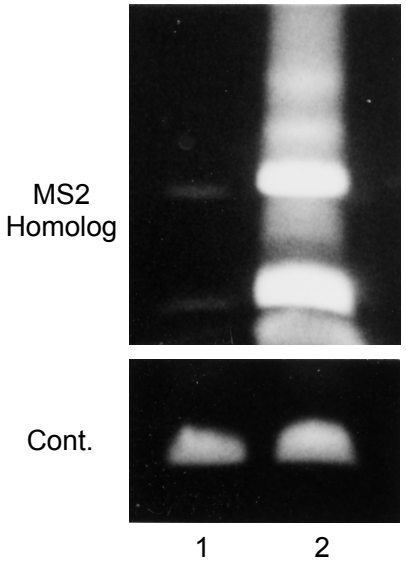

Fig. 1. Transcription of MS2 homolog in *Paramecium tetraurelia*. mRNA samples (1 μg each) derived from preautogamous stock 51 (lane 1) and d4-SL4 (lane 2) were separated by electrophoresis and subjected to Northern blot analysis as described (Tanabe et al., 2002). The resulting blot was hybridized with a cDNA probe specific to MS2 homolog (upper panel) or control for RNA load (lower panel). The control gene has proven to be expressed in an autogamy-independent manner (Tanabe et al., 2003). Chemiluminescent signals from the labeled probe were photographed with an ECL camera.

Several lines of evidence (Tanabe et al., 2003) indicated that MS2 is probably expressed as non-coding RNA (Morey and Avner, 2004; Hutterhofer et al., 2005). Important roles played by non-coding RNA in fundamental cellular processes are becoming clearer, ranging from genetic data storage to sensing, transport, targeting and even regulation of essential events such as cell differentiation and cell death. Among its various types, longer non-coding RNAs, sometimes referred to as mRNA-like non-coding RNA, form a quite distinct class: they are metabolized like mRNA except that significant open reading frames are lacking. Such unique RNA species have so far been known to be involved in X chromosome dosage compensation, genomic imprinting, stress response, etc. (Szymanski et al., 2003) and the MS2 RNA is likely to fall into this group. Most non-coding RNAs exert their function through assuming a distinct base-paired secondary structure with which cognate cellular factors interact specifically. Functional RNA elements (FRE) common to homologous non-coding RNAs are conserved in the secondary structure rather than in the nucleotide sequence, and hence can be hardly detected by primary sequence analysis alone, which may explain the apparent lack of sequence similarity in MS2 to any entry in primary sequence databases (Tanabe et al., 2002). Although identifying phylogenetic correspondents of MS2 is needed to extrapolate its function to other genera, standard sequence alignments are unreliable for this purpose. One promising approach to address this problem is to isolate close homologs of MS2 in ciliates and align them with MS2 for conserved RNA high-

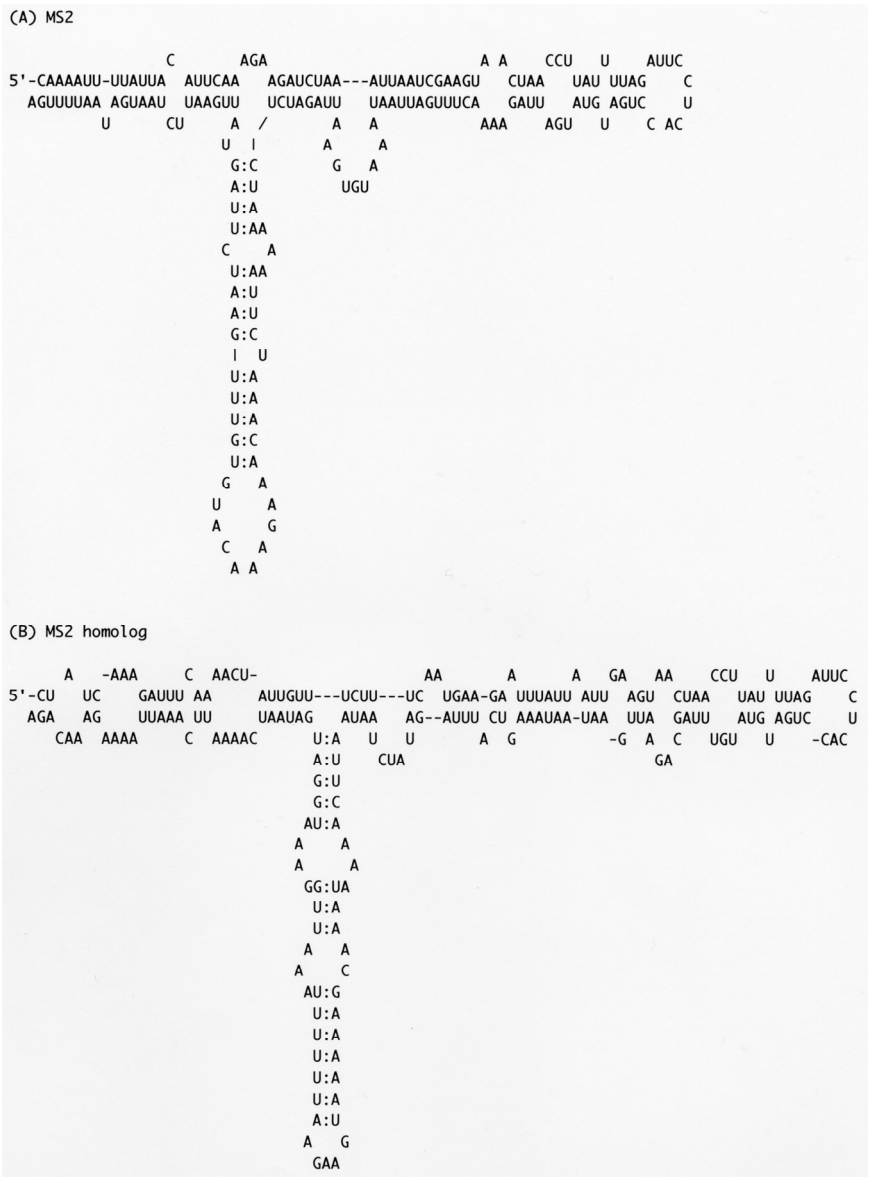

Fig. 2. Representative common structural pattern present in MS2 and its homolog RNAs of *Paramecium tetraurelia.* One of the uniquely folded RNA structures, Type 4, observed in both MS2 (A: 1945-2126 relative to the transcriptional initiation site) and its homolog (B: 1952-2162) were depicted with the RNA folding programs referred to in the text.

ordered structures, through which more distantly related parallels can be identified.

Recently the whole genome sequencing project of *Paramecium* (Dessen et al., 2001; Zagulski et al., 2004) has been completed and the genomic sequence data from *P. tetraurelia* stock d4.2, a derivative of stock 51, suggested the presence of a homolog of MS2 (E. Meyer, personal communication). On the basis of its sequence information, in this study we cloned the full-length cDNA of the homolog (4,106 bp) from d4-SL4 and sequenced it according to the procedure described before

Table 1. List of potential structural homologs to MS2 RNA of *Paramecium tetraurelia*

Folding pattern[a]	Origin	Accession No.	Sequence region[b]
Type 1	*Danio rerio*	AC144824	92,283-92,943
	Danio rerio	AL772373	132,954-133,166
	Danio rerio	BX248081	75,223-75,424
	Danio rerio	BX571727	78,295-78,507
Type 3	*Mus musculus*	AC114425	16,812-16,936
	Danio rerio	AL928714	1,001-1,120
	Danio rerio	AL954739	63,491-63,607
	Danio rerio	BX510913	124,818-124,937
Type 4	*Didelphis verginiana*	AC129066	101,341-101,533
	Bos Taurus	AY644517	75,753-75,970
	Canis familiaris	AC129099	57,614-57,818
	Canis familiaris	AC144643	58,680-58,896
	Mus musculus	AC104396	36,124-36,334
	Danio rerio	AL627262	59,095-59,287
	Danio rerio	AL929543	105,761-105,986
	Danio rerio	BX005271	20,857-21,055
	Danio rerio	BX088526	42,770-42,969
	Danio rerio	CR812483	6,759-6,938

[a]See the text for explanation.
[b]Nucleotide positions are derived from the sequence data deposited in DNA databases.

(Tanabe et al., 2002). The sequence data will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession number AB231862. The homolog was 76% identical to the transcribed region of MS2 (4,044 bp) at the overall nucleotide level. We then investigated the expression of the homolog and found that it showed the same expression pattern as that (Tanabe et al., 2002, 2003) of MS2: much higher expression level in the short-lived mutant than its wild-type parent (Fig. 1) as well as an increasing expression during autogamy (data not shown). Taken together, we concluded that this homolog would serve for the phylogenetically comparative analysis mentioned above.

FRE in non-coding RNA are often correlated with distinct RNA structures because FRE have to fold into the specific conformations in which cell factors can recognize and interact with them. Recent computational experiments have indicated that the structures of FRE are both significantly more ordered and thermodynamically stable than expected by chance (Schultes et al., 1999; Le et al., 2002). The specific structural motif characterized by unique combinations of base pairings and conserved loops is anticipated to be rare in the conformation space formed from a population of the related randomly shuffled sequences. In this study, we first searched for the well-ordered RNA folding in the sequences of MS2 RNA and its homolog by the programs EDscan (Le et al., 2003a) and SigED (Le et al., 2003b) and also searched for unusual folding regions by the SegFold program including SigStb (Le et al., 1988, 1990, 2001). Our computational analyses detected four well-ordered folding patterns consisting of a significant Y-shaped stem-loop structure that were common to the two homologous RNAs. These were located in MS2 RNA at 809-1,004 (designated as Type 1), 1,156-1,256 (Type 2), 1,350-1,458 (Type 3) and 1,945-2,126 (Type 4) counting from the transcriptional start point, which corresponded in the homolog to 822-

985, 1,139-1,244, 1,336-1,444 and 1,952-2,162, respectively (see Fig. 2 for Type 4 which is the most significant consensus structure). These structured elements were not superposed on regions in which the two RNAs share the highest sequence identity, which is diagnostic of functional RNAs as stated above. Based on the results of this structural comparison, we searched for structures homologous to the four patterns by scanning genomic and cDNA sequences of other genera using the programs rna_match (Collins et al., 2001) and HomoStRscan (Le et al., 2004). The results were summarized in Table 1. These unique Y-shaped conformations except Type 2 were found also in several animal species but not in human, *Caenorhabditis elegans*, and *Drosophila melanogaster*. These unusual folding patterns are likely to be present in transcription products which remain yet to be found out in many organisms. No biological information and features about these potential FREs on the nucleotide sequence level were found in the genomic regions deposited in the DDBJ/EMBL/GenBank databases. We further performed database searches against entries in RNAdb (Pang et al., 2005), a mammalian non-coding RNA database, for the presence of annotated secondary structure signals in the mammalian structural homologs detected here, but at present no significant hit has been obtained for any homolog sequence. Collection of further experimental data and the following refinement of predicted RNA structure as well as future enrichment of RNA databases would help reveal more extensive phylogenetic conservation and biological roles of the folding structures found in this study. Although functional characterization of the structural features observed in the *Paramecium* RNAs and their potential structural homologs requires more detailed experimental and computational works, the present result is presumably an important step toward understanding the life-programming mechanism that holds in all of the organismic species with an intrinsic upper limit of longevity.

ACKNOWLEDGMENTS

We thank Dr. Eric Meyer for his valuable suggestions. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

REFERENCES

Collins, G.D., Le, S-Y. and Zhang, K. (2001) A new algorithm for computing similarity between RNA structures. Information Sciences, 139, 59-77.

de Magalhaes, J.P. (2005) Open-minded scepticism: inferring the causal mechanisms of human ageing from genetic perturbations. Aging Res. Rev., 4, 1-22.

Dessen, P., Zagulski, M., Gromadka, R., Plattner, H., Kissmehl, R., Meyer, E., Betermier, M., Schultz, J.E., Linder, J.U., Pesrlman, R.E., Kung, C., Forney, J., Satir, B.H., Van Houten, J.L., Keller, A-M., Froissard, M., Sperling, L. and Cohen, J. (2001) *Paramecium* genome survey: a pilot project. Trends Genet., 17, 306-308.

Guarente, L. (2005) Calorie restriction and SIR2 genes-Towards a mechanism. Mech. Aging Dev., 126, 923-928.

Hutterhofer, A., Schattner, P. and Polacek N. (2005) Non-coding RNAs: hope or hype? Trends Genet., 21, 289-297.

Le, S-Y., Chen, J.H., Braun, M.J., Gonda, M.A. and Maizel, J.V. (1988) Stability of RNA stem-loop structure and distribution of non-random structure in the human immunodeficiency virus (HIV-I). Nucleic Acids Res., 16, 5153-5168.

Le, S-Y., Chen, J.H. and Maizel, J.V. Jr. (1990) Efficient searches for unusual folding regions in RNA sequences. *In:* Structure & Methods: Human Genome Initiative and DNA Recombina-

tion. Sarma, R.H. and Sarma, M.H. (eds.). Adenine Press, Schenectady, pp. 127-136.

Le, S-Y., Liu, W.M., Chen, J.H. and Maizel, J.V. Jr. (2001) Local thermodynamic stability scores are well represented by a non-central student's t distribution. J. Theor. Biol., 210, 411-423.

Le, S-Y., Zhang, K. and Maizel, J.V. Jr. (2002) RNA molecules with structure dependent functions are uniquely folded. Nucleic Acids Res., 30, 3574-3582.

Le, S-Y., Chen, J.H., Konings, D. and Maizel, J.V. Jr. (2003a) Discovering well-ordered folding patterns in nucleotide sequences. Bioinformatics, 19, 354-361.

Le, S-Y., Chen, J.H. and Maizel, J.V. Jr. (2003b) Statistical inference for well-ordered structures in nucleotide sequences. Proceedings of the 2003 IEEE Bioinformatics Conference (CSB2003), Stanford, CA, 190-196.

Le, S-Y., Maizel, J.V. Jr. and Zhang, K. (2004) An algorithm for detecting homologues of known structured RNAs in genomes. Proceedings of the 2004 IEEE Bioinformatics Conference (CSB2004), Stanford, CA, 300-310.

Morey, C. and Avner, R. (2004) Employment opportunities for non-coding RNAs. FEBS Lett., 567, 27-34.

Pang, K.C., Stephen, S., Engstrom, P.G., Tajul-Arifin, K., Chen, W., Wahlestedt, C., Lenhard B., Hayashizaki, Y. and Mattick, J.S. (2005) RNAdb—a comprehensive mammalian noncoding RNA database. Nucleic Acids Res., 33, D125-D130.

Schultes, E.A., Harber, P.T. and Labean, T.H. (1999) Estimating the contribution of selection and self-organization in RNA secondary structure. J. Mol. Evol., 49, 76-83.

Sonneborn, T.M. (1954) The relation of autogamy to senescence and rejuvenescence in *Paramecium aurelia*. J. Protozool., 1, 38-53.

Sonneborn, T.M. (1974) *Paramecium tetraurelia*. *In:* Handbook of Genetics Vol. 2. Mayer, E. (ed.). Plenum, New York/London, pp. 469-594.

Szymanski, M., Erdmann, V.A. and Barciszewski, J. (2003) Noncoding regulatory RNAs database. Nucleic Acids Res., 31, 429-431.

Takagi, Y., Suzuki, T. and Shimada, C. (1987) Isolation of a *Paramecium tetraurelia* mutant with short clonal life span and with novel life features. Zool. Sci., 4, 73-80.

Takagi, Y., Izumi, K., Kinoshita, H., Yamada, T., Kaji, K. and Tanabe, H. (1989) Identification of a gene that shortens clonal life span of *Paramecium tetraurelia*. Genetics, 123, 749-754.

Takagi, Y. (1999) Clonal life cycle of *Paramecium* in the context of evolutionally acquired mortality. In: Progress in Molecular and Subcellular Biology Vol. 24. Macieira-Coelho, A. (ed.). Springer-Verlag, Berlin/Heidelberg, pp. 81-101.

Tanabe, H., Fujiike, M. and Yoshioka, S. (2002) *Paramecium tetraurelia* genes expressed in an age-dependent manner. Jpn. J. Protozool., 35, 143-150.

Tanabe, H., Fujiike, M. and Yoshioka, S. (2003) Isolation and characterization of genes showing altered expression specific autogamy in *Paramecium tetraurelia*. Jpn. J. Protozool., 36, 1-8.

Zagulski, M., Nowak, J.K., Mouel, A.L., Nowacki, M., Migdalski, A., Gromadka, R., Noel, B., Blanc, I., Dessen, P., Wincker, P., Keller, A-M., Cohen, J., Meyer, E., and Sperling, L. (2004) High coding density on the largest *Paramecium tetraurelia* somatic chromosome. Curr. Biol., 14, 1397-1404.